Universidad Nacional Mayor de San Marcos

Decana de América

DOCUMENTO DE TRABAJO
N° 2022-02

Instituto de Investigaciones Económicas

NOTES ON ECONOMETRIC METHODS

Mario D. Tello

Documento de Trabajo N° 2022-02

Editado por el Instituto de Investigaciones Económicas
Facultad de Ciencias Económicas, UNMSM
Av. Carlos Germán Amezaga #375 -Cercado de Lima – Perú
investigacion.economia@unmsm.edu.pe

NOTES ON ECONOMETRIC METHODS

Mario D. Tello

Febrero, 2022

ENLACE DE LA PUBLICACIÓN

NOTES ON ECONOMETRIC METHODS

Mario D. Tello[1]

[1] Professor at the Economic Sciences Faculty of the Universidad Nacional Mayor de San Marcos.

PREFACE

These notes have been developed whereas as I was a Visiting Professor at the International Development Program (IDV) at University of Southern Mississippi, College of International & Continuing Education, International Development Doctoral Program USM in the fall semester of 2000. These notes are not a substitute for any graduate textbook of Econometrics nor are complementary. Instead, these notes reflect the methods that most second year students need for their thesis. As the program evolves other research techniques will be used and some additional notes will be required. The notes are a brief overview of some of the standard techniques in econometrics. In that sense, these notes are incomplete. Students of the IDV may find useful in the sense, that it provides a glance of the econometric methods they need to read, study, and understand in order to be able to deal with their doctoral thesis projects. The notes are useful for analyzing issues such as:

1. The economic growth impact of human capital; institutions, the gaming sector; cultural and gender factors, and expenditures on peacekeeping missions.

2. The factors that determines employment and wages within a determined workforce such as immigrants.

3. The factors that determines the demand of new technology and its diffusion. Examples are electronic commerce, broadband services, and environmental/cleaner technologies.

These issues constitute the ongoing research work of the first-class students of doctoral students at any university. The notes are easy to read and use low as well as a medium level of mathematics. To understand better the notes, it is suggested that students read some basics on Matrix Algebra and Calculus, Mathematical Analysis and Statistics. I owe my gratitude to Professor Mark Miller from the Economic Development department and the dean of College of International and Continuing Education at USM, Tim Hubson for allow me to visit the department and college.

INDEX

Pag

Pag

LIST OF FIGURES

INTRODUCCION

Let me start with a series of examples to provide an idea what this note is all about from my own perspective.

Example No 1:

> *"Looking at Table No 1(using country data from international sources) there are several questions that do arise: i) ¿ why there exist the difference between the level and the rate of growth of per capita GDP among these countries? ii) ¿Is there any relationship between the rate of export growth and per capita GDP? iii) ¿ is it possible for LDCs to achieve the per capita GDP level of USA?"*

Example No 2:

> *"Telecommunications technology is changing to a surprisingly speed, just few decades ago our communications were based upon phones, facsimiles, telegraph and so on. Now, we have the Internet that could be transmitted by cable or by broadband. Table No 2, from an article published in Applied Economics in 1997 show that for a sample of 100 households for each of the six capital states cities from Australia plus 300 households' interviews from small towns, and 100 more households from Canberra."*

This figure reveals that 7% of the sample demanded broadband residential services. ¿ What are the main factors that determines this demand? This is another type of question that needs a research methodology to answer it.

Example No 3:

> *"Figure No 1 drawn from the Panel Study of Income Dynamics of the US show that by 1988 there still existed a 12% gap between the women relative to the men wages. In other, in average men earns 12% more money than women in the US. Table No 3 also confirm this gap between recent graduates compared to college seniors from a sample of 219 recent graduate students and 248 senior college students."*

As in previous examples, several questions do arise from these two data sets. ¿ What factors explains the gap difference between the women and men salary? ¿why these gaps does not disappears through time? Here again we need a research design in order to answer these questions.

A. COMPONENTS OF A RESEARCH DESIGN:

There are at least 9 components:

1. Definition and Relevance of a problem or questions that need to be solve;
2. A set of Data that can be used in order to assess the problem;
3. A model/theory or a set of hypotheses that suppose to explain the problem;
4. A research technique to be used to assess the data and drawn from this a set of results
5. Analysis of the data/information;
6. Implementation of the research techniques;
7. Model/theory/hypotheses Validation;
8. Results and conclusions from components 5-7
9. Deficiencies and Caveats of the research design

Let's take as an example the No1. Here the problem is obvious is how poor countries becomes a rich country. The data are quantitative and qualitative data from poor countries on their degree of development. The models are the set of theories that suppose to explain the development process. You have studied some of those last year. The technique is usually regression analysis where Robert Barro is one of the major contributors.

B. ON THE DATA:

The data can be classified by: type; random features; dimension; location; and time. By type the data could be Quantitative or Qualitative. In the first case, variables can be observed and measurement directly. Example, individual income, expenditures, etc. In second case, we can transform the data to categorized and ordered data. Example: sex, age, and level of education and so on. By random features, data can be experimental or non-experimental. The first one is used by natural scientists. The second is used by social scientists. By dimension, the data could be a set of samples or the whole population or universe. By location we mean geographically. The data could be a region, a city, a country and set of countries. The data could by period of time or a time series data or cross-section data from a predetermined period of time.

The data en Example 1, is quantitative, a time series, non-experimental, and a set of countries. In the second example the data is non-experimental, cross section and a sample of quantitative and qualitative data. En example No 3, of the same nature of the second sample.

C. ON THE MODELS AND/OR HYPOTHESES

These are mathematical or logical ideas about how "the world works in abstract" under very limited and usually designed nature of the world. What are the elements for a model or theory can be considered as paradigm have been studies by philosopher such as Kuhn[2] (1962) [The Structure of Scientific Revolutions] and Lakatos [Falsification and the Methodology of Scientific Research Programs] and we will not look this literature. Rather, the student will use the standard and new theoretical literature of the problem/question that she or he wants to deal with, and this will be used as starting point of the research.

As matter-of-fact regular PhD programs suppose to teach to students the techniques of how to elaborate model, ideas and even a theory.

D. RESEARCH TECHNIQUES

There exists an extensive literature on these techniques and from all social and natural sciences. Economists, use Econometrics, Biologist use Biometrics, sociologists use sociometrics and so on. Practically all these techniques use as its main framework the theory of statistics.

E. RESULTS AND CONCLUSIONS

Arise from the analysis of the data and the application of the research techniques.

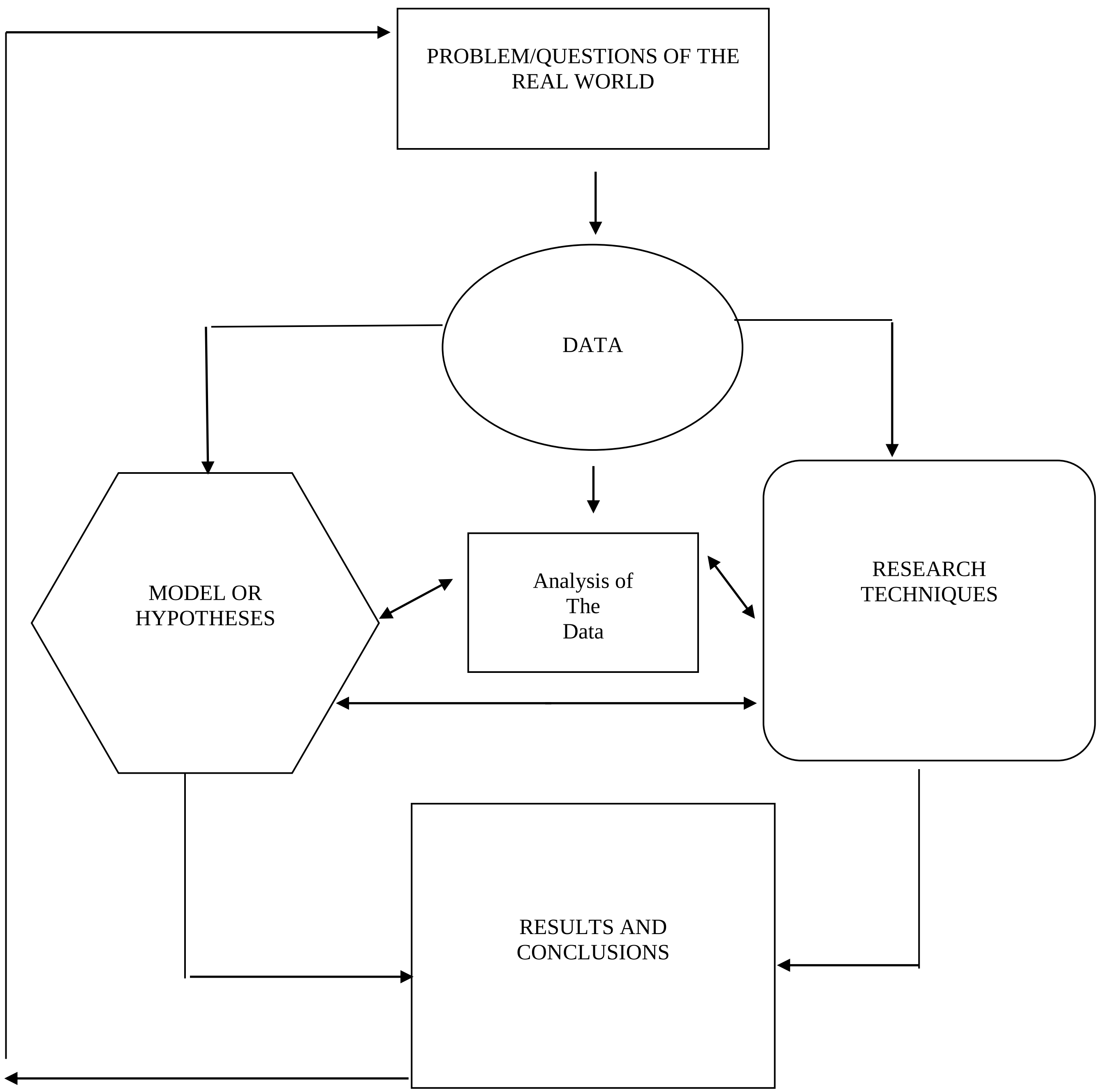
This note will try to do two things. One is to have a first draft of a proposal taking into consideration the components considered here. For that purpose, it is needed the review of the literature of the thesis area or topic. So, it is very important that the selected papers (of that review) be as related as it is possible to the research or dissertation proposal.

The other is to learn some research techniques. We will concentrate upon those are very common in many applications. Those related to econometrics and statistics.

[2] Hunt (1989) [Economic Theories of Development] has a brief comment on these ideas and concepts.

Figure N° 1

Components of a research design



LECTURE No I

INTRODUCTION TO THE METHODOLOGY OF ECONOMETRICS

INTRODUCTION:

The last meeting it was described the components of a research work. These are:

1. Definition and Relevance of a problem or questions that need to be solve;
2. A set of Data that can be used in order to assess the problem;
3. A model/theory or a set of hypotheses that suppose to explain the problem;
4. A research technique to be used in order to assess the data and drawn from this a set of results
5. Analysis of the data/information;
6. Implementation of the research techniques;
7. Model/theory/hypotheses Validation;
8. Results and conclusions from components 5-7
9. Deficiencies and Caveats from the research design

The continuo process of this research design is what is called Science.

According to Hendry (1980):

> *"Science is a public process. It uses systems of concepts called theories to help to interpret and unify observation statement called data. It turn the data are used to check or "test" a theory, [to predict future observations and to solve real problems]".* pp. 388.

Next semester we will spend more time in two components 2 and 3 [the data and the theory]. What is remaining in this course, we will be devoted to the basic understanding of some research techniques. Specifically, econometric techniques or methods. Before we get into this discussion, we need to spend some time to discuss the *nature of the data that a social science researcher faces* and its shortcomings.

Let me provide some examples to show the issues. I will use the same set of data/examples from the first meeting.

Problem 1: ¿ Does Exports Growth leads to Economic Growth?

Problem 2: ¿ what factors restrain the demand for broadband services?

Problem 3: ¿ are there wages discrimination against women in the US?

In order to address, these questions we need not theory but also data from the real world. Once we have the data, we need some methodology to work with the data in order to draw inferences from it. This methodology must be a serious and formal, since often the results of the research will affect the lives of million of people.

I will use Problem 1 to understand better the issues involved and because I have worked a long time in that area. Further, nowdays International organizations such as the World bank and the International Monetary Funds are applying or suggesting policies to developing countries based on inferences from the data dealing with Problem No 1. What it has been found for many researchers, including myself, is that the huge empirical literature that these organizations based upon their recommendations have serious shortcomings and no necessary are solid and close to what is happening is developing countries. To illustrate my comments, suppose we have the following data set:

	Rate of Growth of Per capita GDP	**Rate of Growth of Exports**
	Y	X
	2	2
	1	3
	9	10
Mean (ym,xm)	4	5

yd= yi-ym; xd=xi-xm;

Sum	yd	Σyd^2	yd.xd	xd	Σxd^2
Deviations	-2	4	6	-3	9
	-3	9	6	-2	4
	5	25	25	5	25
Sum		41	37		41

$sy= [41/3]^{1/2}$; $sx= [41/3]^{1/2}$

$rxy= [(\Sigma xd.yd)/n]/sx.sy= [37/3]/[41/3]= 37/41= 0.902439$

Two interpretations of these results are:

1: There is a positive association between both rate of growth
2: This association may be significant[3]

From this, ¿can we conclude that:

1. The real export is the engine of growth in developing countries; or that
2. The policies that LDCs needs to implement are liberal policies since these policies will increase exports?

Despite of the statistical problems international institutions have concluded these two last statements.

From this data and results: we cannot conclude anything, statistically even less for policy recommendation. Let come back to the data. The first two rows the data seems to indicate contrarily a negative rather than a positive association. The reason why the association is positive and statistically significant is because the large numbers (relative to the other two) of the third observation (this is an outlier problem). Further, there is not enough data. Not only the policy are wrong since the data are not saying that but also the data does not say anything about causality nor ¿how do we know that a liberal market do in fact increase exports? This little example, show the difficulty to draw inferences from the data and the importance and providing policy recommendations from the data. We need to be serious, rigorous and cautious.

1. BASIC CONCEPTS:

¿What is the nature of the data that faces social scientist? Let me provide a simple:

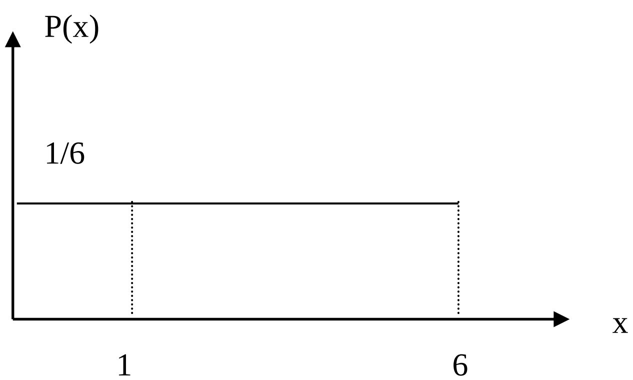
Throw a dice with six options which are defined by S= {1,2, 3, 4, 5, 6}. If this dice is fair (unloaded) then the probability to get any particular number is P(x=a) =1/6. What are the components of this example:

1.1 All the possible elements of this event, we called that S. Then S is called population or the whole Universe of the event;

1.2 Let's call x to any of the six elements of the event, x is called a random or aleatory (aleatoric) variable;

[3] From Kendall (1952): the correlation coefficient test is the following: Ho: $\rho=0$; H1: Not Ho; the statistic used in the t-student with N-2 degrees of freedom, where $t= rxy/[(1-rxy^2)/(N-2)]^{1/2} = 0.389$. This coefficient is not significant statistically.

1.3 The probability of the frequency of any outcome is called the probability (frequency) function/distribution of x. In this case $P(x)= 1/6$ for any x that belongs to S. Any event that has these components is called Uncertain/Risky Events. The name random is due to the fact that the event is uncertain and the fact that it is feasible to represent this uncertain event through a probability or frequency function. This example shows the role of the statistics in analyzing the data, in particular data from uncertain events. The statistics teach us to organize those uncertain or aleatory elements in a way that we can deal with in a technical way. Moreover, allows saying something about those uncertain events. Now the Event of the throwing a dice can be able to summarize through the probability function:



Some features of this probability are:

Population Mean/First Moment/Expected Value of x= $\sum_{i=1}^{n} xi.P(xi) =E(x)=\mu$

Population Variance/Second Moment relative to the mean of x = $\sum_{i=1}^{n} (xi- E(xi))^2. P(xi)$ =

$E[x- E(x)]^2 = \sigma^2$;

It should be noticed that both the mean and the variance of the universe are no longer random variables. They are fixed numbers called the Parameters of the distribution/probability of x. In the dice example, $\mu= 3.5$; = $\sigma^2 = 3.75$; $\sigma=1.94$. Now x could be a discrete or continuos variable. In the former case, the probability is called a density function. In the latter, the function is called a probability function. P(x) or f(x) are defined to be a nonnegative values and numbers between 0 and 1. Further the sum of the probability for the whole universe must be one. In math terms:

$0\leq P(x) [f(x)] \leq 1$; and $\Sigma P(xi)=1$; or $\int f(x)dx=1$

x could be timeless or in determined space or could be through time. In the case that xt be through time, the random variable is called a random process. Finally, we can analyze one variable (i.e., a univariate analysis) or several variables (i.e., a multivariate analysis).

Let's provide now a more interesting example. Let x is the monthly gas expenditure of a particular household. S is the set of possible values of x.

S= {200, 400, 100, 250, 500, 100, 200, 400, 300, 444, 500, 600}

Question: Is x from S a random variable? Yes, no and why?

Yes, it is a random variable. The reason is because of there are many and uncertain sources that explains the changes of x:

i) Price changes; households do not know when this event will occur;
ii) Quantity change;
iii) Weather conditions;
iv) Any unforeseen event (accidents, robbery, fire, etc.)

Even if there is a way to control for price and for unforeseen events, x is random because the human nature of the decisions and the weather element.

Since social sciences deal with events, that ultimate are related or caused by human behavior then the nature of the data is that they are random variables. This is the major problem with the data. For this nature, never we can know the true. We can have an approximation in terms of probabilities. Statistics help us to manipulate with those kinds of variables otherwise we could not say anything from the data.

2. EXPERIMENTAL VS NON EXPERIMENTAL DATA

One of the major criticisms against the research techniques in social sciences is the fact that the data does not come from a controlled experiment. Usually in natural sciences the researcher controls the variables that they want to associate. Let's take the Leamer (1983) example:

A researcher wants to investigate the effect of a fertilizer in the yield/output of a particular agricultural commodity, lets say potato. So he assigned in a random fashion the fertilizer in a set of plots. After a certain period of time, he can compare the average yield/output of the plots with fertilizer with the average yield/output of the plots without the fertilizer. From this comparison, he could find out the effect of the fertilizer on the yield. This is an experimental random sample.

The social sciences researchers/economist do not have this kind of sample. In his case, the researcher has a set of data from yields of some set of plots which some of them has been exposed to trees where birds roost and produce birds droppings. In this example, bird droppings are assumed to be the fertilizer.

Both cases (the experimental and non-experimental examples) however, share a common element: this is RANDOMIZATION. Whereas, in the experimental data, the research produce the sample in random way and controlling some other factors [weather, trees, etc] in the non-experimental case we have a data where the nature (or human behavior) does the randomization. A random experiment only means that the data is probable to be adequately mixed but does not assure that all the data are exactly mixed. Moreover, it could happen that one of the infinite outcomes of the randomization of the experiment be just equal to the ones collected by the social science researcher. As long as, both types of experiment are random then the difference is just of degree. The non-experimental data usually will have larger errors than the errors from experimental data. If experimental science do control for unimportant factors, by the same token it is possible to control variables in the non-experimental case. In order to do that, however, the social science researcher need to make some assumptions about the data that allow him to control for other factors. These assumptions according to Leamer (1983) makes econometrics (social sciences techniques) whimsy and fragile. Pratt and Schlaifer (1984) present the same kind of arguments.

3. MEASUREMENT WITHOUT THEORY/CAUSALITY AND THE ROLE OF THE THEORY IN THE ECONOMETRICS METHODOLOGY

¿How the Econometric Methodology, given the data limitations, may infer relevant propositions from the data? In the export/growth example, the data show that there is a positive association between the rate of growth of export and per capita GDP, from here ¿Can we say that Export growth do cause per capita GDP growth and this effect is positive? ¿Are the data information enough for making causality inferences among variables? In order to answer these questions, we need first to address two issues. One is the role of the theory and the other is the definition of causality.

There are several reasons why measurement without theory it is not enough for solving the problems that social science faces. The main arguments presented here are drawn upon T. Koopmans (1947) who critics the data analysis made by Burns and Mitchell (1946, Measuring Business Cycles, NBER.).

> **A1:** *"[Data Analysis without theory] are concerned with the features of the data rather than the underlying behavior of man. [Data, at the end is one way that the human behavior can be represented]".* pp 162. Koopmans (1947)

A2: *"Even for the purpose of systematic and large scale observation, theoretical preconceptions about its nature cannot be dispensed with"* pp. 163. Koopmans (1947).

Under a Theory or model the researcher knows what variables to select and what are the relevant thus Social Science may explain events.

A3: *"Economist are not in position to perform experiments with an economic system as a whole for the sole purpose of establishing scientific truth. It is not possible in many economic problems to separate causes and effects by varying causes one at time, studying the separate effects-a method so fruitful in natural sciences….. Economic theories are based upon a different kind of evidence [from natural sciences] such as knowledge of the motives and habits of consumers and of the profit-making objectives of business enterprises….. Although the economic theories are incomplete and in need of reformulation and elaboration such theory as we have is an indispensable element in understanding in a quantitative way the formation of economic variables. [Measurement is needed in order to identify the structural equations given by the theory. The reason of this is because structural equations or theories] are needed to make relevant conclusions for economic policies".* pp. 167. Koopmans (1947).

A4: *"In dynamic economies the phenomenon itself is either essentially a stochastic [random] process or needs to be treated as such because of the great numbers of factors at work. The main problem of inference is the choice of the statistics [defined as a function of the observations]. In order to make the appropriate inference we need also the theory, In order that a theory be useful with need appropriate data and statistical inference".*

A5: *"According to Feigl (1953, "Note on Causality". H. Feigl and M. Brodbeck. Eds. Readings in the Philosophy of Sciences] a causality means predictability according to a law or a set of laws"*. pp 7. Zellner (1988)

Summing up, the roles of the theory are i) to organize and select the relevant data to explain an event in the social sciences ii) for policies recommendations; iii) to define causality. Data is needed in order to test the relevance of the theory and learn more about the event.

4. ECONOMETRICS AND THE ECONOMETRIC TECHNIQUE METHODOLOGY:

D1: *"Econometrics is the advancement of economic theory in its relation to statistics and mathematics"* [Econometric Society founded in 1930];
D2: *"Econometrics is the study of the properties of data generation processes, of techniques for analyzing economic data, methods of estimating numerical magnitudes of parameters with unknown values, and procedures for testing economic hypotheses: it*

plays an analogous role in primarily non-experimental disciplines to that to the statistical theory in inexact experimental sciences”. pp. 591, Hendry (1980).

Figure No 2 illustrate the classical econometric methodology

4.1 We need the theory/model/ hypotheses, which is defined by a mathematical economic model:

[1] Yit= F(Xit, Yit; Λit); F is a mathematical function

Yit is a vector of g variables that are called endogenous variables, which are the variables that the model/theory/hypothesis try to explain;

Xit is a vector of k variables that are called exogenous variables, which are the variables that the theory/model/hypothesis use in order to explain Yit;

i is the observation of an economic agent; t is the time observation of the economic agent i;

Λit is matrix of size gxk of unknown Parameters that defines F;

4.2 We need the statistical theory to be able to draw inferences from a non-experimental data set.

With the statistical theory the former equation is transformed to:

[2] Yit= F(Xit, Yit, Λit, Uit);

Where Uit is the vector of g stochastic terms of variables that are called the Error or stochastic term. In order to make meaningful inferences from the data we need certain assumptions on Uit, and F. Equation [2] are called the structural stochastic equations, the structure or the stochastic economic laws that suppose to explain the set of data.

Using this framework, the econometric technique obtains statistical and formalized methods to estimate Λ using statistics or estimator and makes statistical inferences on these estimators. Let’s call Λe (Xit, Yit, Uit) be the estimators that depend upon the data set and the stochastic term. Λe is a set of random variables because depends upon Uit.

¿How many estimations methods there are? They are infinite, but those with sound statistical properties only a few. Some of these few are the methods that students at the IDV will use in their dissertation topic. The next chapters provide a brief overview of some these methods.

REFERENCES

Porier, D., 1994, The Methodology of Econometrics. Vol. No 1, 2. The International Library of Critical Writings in Econometrics, No 6. Elgar Reference Collection, Cambridge, England.

Black, F., 1982, “The Trouble with Econometric Models”. *Financial Analyst Journal*, March/April, 35, pp. 3-11.

Hendry, D., 1980, “Econometrics-Alchemy or Science?”. *Economica,* November, 47, pp. 387-406. (Section I, II).

Kendall, M., 1952, The Advanced Theory of Statistics, Hafner Publishing Co., New York

Keynes, J, M., 1939, “Professors Tinberger’s Method”. *Economic Journal*, XLIX, pp. 558-568.

Koopmans, T., 1947, “Measurement Without Theory”. *Review of Economic and Statistics*, 29, pp. 161-172.

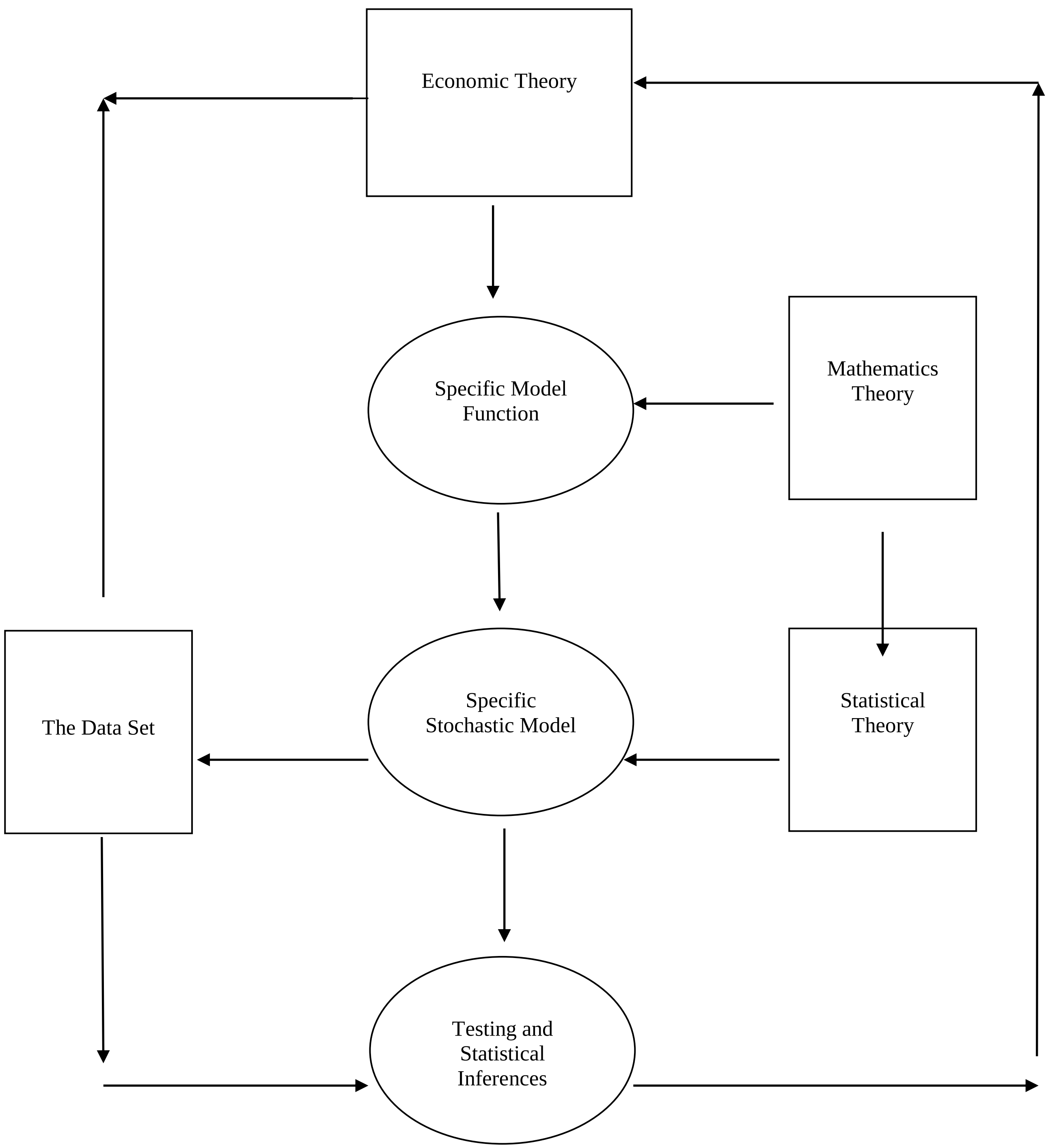
Kuhn, T.S., 1962, The Structure of Scientific Revolutions, Chicago University Press.

Lakatos, I., 1974, “Falsification and the Methodology of Scientific Research Programmes”. In I. Lakatos and A. E. Musgrave, Criticism and the Growth for Knowledge, pp. 91-196. Cambridge University Press.

Leamer, E., 1983, “Let’s Take the Con Out of Econometrics”. *American Economic Review*, LXXIII, pp. 31-43.

Figure N° 2

The classical econometric methodology



Poirier, D., 1994, "Introduction".

Popper, K., 1968, The Logic of Scientific Discovery. London: Hutchinson.

Pratt, W., R. Schalaifer, 1984, "On the Nature and Discovery of Structure". *Journal of the American Statistical Association*, 79, No 385, pp. 9-21.

Tinbergen, J., 1940, " On A Method of Statistical Business-Cycle Research: A Reply". *Economic Journal*, L, pp. 141-154.

Tintner, G., 1953, "The Definition of Econometrics". *Econometrica*, 21, pp. 31-40.

Wold, H., 1969, "Econometrics as Pioneering in Non Experimental Model Building". *Econometrica*, 37, No 3, July, pp. 369-381.

Zellner. A., 1988., " Causality and Causal Laws in Economics". *Journal of Econometrics*, 39, No ½, September/October, pp. 7-21.

LECTURE No II

BASIC CONCEPTS: RANDOM VARIABLES AND PROBABILITY THEORY

A random variable is a particular feature of a uncertain event. If x is a random variable then there exist a density/probability/frequency function associated with it. Any function f (x) is a density/probability/frequency function if it satisfies the following conditions:

1. $0 \leq f(x) \leq 1$; 2. $\sum_{i \, \varepsilon \, S} f(xi) = 1$ for discrete variables and $\int_{x \, \varepsilon \, S} f(x)\, dx=1$

Where S is the Population /Universe/Whole Domain of x.

Accumulative Function is $F(x=a)= P(xi \leq 0)$; or and $\int_{-\infty}^{a} f(x)\, dx=1$

Univariate random variable is one random variable. Multivariate random variables is a set of random variables X ={ x1, x2,….xn} defined by the Joint Probability distribution F(X). For practical reasons we will deal with continuos random variables.

Marginal Density Function $f(xi)= \int \int \ldots\int f(xi,\ldots ,xj)\, dx1.dx2..dxj$; where xj is a random variable different than the xi random variable;

Conditional Density Function $f(Xi/Xj)= f(Xi, Xj)/f(Xj)$; where Xi, Xj are set of different random variables;

xi and xj are independent random variables if $f(xi, xj)=f(xi).\, f(xj)$ for any i,j. It means that these variables are not related in any way.

Let F(X) a joint distribution. The set of parameters Λ defines the features of this distribution. Some of these features are called the Moment.

$E(x^r)= \int x^r f(x).dx$; is the r moment of f(x);

If r=1, is called the first moment or the population mean represented usually by μ;
If r=2, is called the second moment;

A compose moment is when:

$E(G(x)) = \int G(x).\ f(x).dx$; if $G(x) = x - E(x)$ then we have the moments with respect to the first moment. That is:

$E[(x - E(x)]^r = \int [x-E(x)]^r\ f(x).dx$; is the r moment with respect of E(x) of f(x);

If r=2, then we have the second moment with respect to its mean. This is called the ***variance*** of x represented usually by $\boldsymbol{\sigma^2}$.

Covariance (xi, xj)= E[(xi-E(xi).(xj-E(xj))]/ $[V(xi).\ V(xj)]^{1/2}$. It measures the degree of linear association between xi and xj random variables.

A Random Sample; is a set of values for a random variable or a set of random variables. For any pair of random variables of the random sample, they are independent each other and each has the same density function;

If two random variables are independent then cov(xi, xj)=0;

If two random variables are linearly independent the cov (xi, xj)=0;

If two variables are linearly independent that does not mean that they mutually independent;

An estimator/statistic is a function of the random sample of size n. Λe= H(x1, x2,..xn)

REFERENCES

Any textbook of Statistic.

Anderson, T., 1958, An Introduction to Multivariate Statistical Analysis, Wiley.

Cox, D., D. Hinkley, 1974, Theoretical Statistics, London: Chapman and Hall.

Creighton, J., 1994, A First Course in Probability Models and Statistical Inference.

Kendall, M., 1952, The Advanced Theory of Statistics, Hafner Publishing Co., New York.

Rao, C, 1965, Linear Statistical Inference and Its Applications, Wiley Springer-Verlag, New York].

Any textbook of Econometrics (the Statistical Appendix).

Dhrymes, P., 1974, Econometrics Statistical Foundations and Applications, Harper & Row Publishers.

Johnston, J., 1980, Econometric Methods, McGraw-Hill.

Judge, G., R. Carter, and associates, 1995, The Theory and Practice of Econometrics, John Wiley & Sons.

Maddala, G., 1990, Econometrics, McGraw-Hill].

LECTURE No III

BASIC DESCRIPTIVE STATISTICS

Give a random sample of size n, $S=\{x1, x2, x3 \ldots xn\}$ we can define the following set of statistics or estimators:

1. Sample Average/Mean $xm = \mu e = \sum_{i=1}^{n} xi/n$;

2. Sample Geometric mean $xg = [\Pi xi]^{1/n}$;

3. Sample Variance $s^2 = \sum_{i=1}^{n} (x-xm)^2 /n$;

4. Unbiased sample variance $s1^2 = \sum_{i=1}^{n} (x-xm)^2 /(n-1)$;

5. Population Simple Correlation coefficient between xi and xj (two different random variables) $\rho = cov(xi, xj)/(\sigma xi.\sigma xj)$;

It can be proved that $-1 \leq \rho \leq 1$; ρ measure the degree of linear relationship of two random variables. The closer is $/\rho/$ to 1 the higher is the degree of linear relationship if $/\rho/ =1$ then xj or xi can be written as $\alpha.\ xi + \beta.xj = \gamma$; where α,β, and γ are parameters. The next definitions are appropriate when there are two or more random variables. Thus, we will assume that we have a multivariate random sample $S=\{ (xi, xj, xk)\}$ of at least three random variables. The size of the sample of each random variable is n.

6. Sample correlation coefficient: $r= \Sigma\ (xi-xmi).(xj-xmj)/[\Sigma(xi-xmi)^2.\ \Sigma(xj-xmj)^2]^{1/2}$;

7. Population Partial Correlation Coefficient between two variables xi, xj conditional to the xk variable $\rho p= cov(xi,xj/xk)/ /(\sigma xi/xk.\sigma xj/xk)$;

Partial correlation coefficient is defined on conditional distribution, and simple correlation coefficient is defined under joint or unconditional distribution.

8. Sample Partial Correlation
Coefficient between two variables $rpij.k= -(/Rij/)/[/Rii/./Rjj/]^{1/2}$;
xi, xj conditional to the xk variable

Let be R the square matrix of order k (which is number of random variables) then:

$$
R= \begin{pmatrix} r11\ r12\ \ r1j\ ...r1k \\ ri1\ \ ri2\dots\ rij\ ....rik \\ \\ rk1\ rk2\qquad\quad rkk \end{pmatrix}
$$

rij are the sample correlation coefficients of any pair of random variables. /R/ is the determinant of that matrix and /Rij/ is the cofactor (i,j) of the matrix R. The cofactor is the determinant of the matrix that arise from deleting row i and column j of matrix R. The sign of the cofactor is $(-1)^{i+j}$.

Proportions are also random variables which probability could be a binomial or a Bernoulli Distributions (see below). Examples: i) the probability that a particular candidate wins in a political race; ii) the probability some particular event of interest happens.

Contingency Tables are tables that summarize the different outcome/features of two or more random variables. Example:

		Variable Xj					
		X1	X2	………	Xj	Xm	
VARIABLE Yi	Y1						
	Y2						
	. .						
	YN						

In this table of two random variables Xj and Yi, where j=1, m ; and i=1,..n .

REFERENCES

Any textbook of Statistic.

Anderson, T., 1958, An Introduction to Multivariate Statistical Analysis, Wiley.

Cox, D., D. Hinkley, 1974, Theoretical Statistics, London: Chapman and Hall.

Creighton, J., 1994, A First Course in Probability Models and Statistical Inference.

Kendall, M., 1952, The Advanced Theory of Statistics, Hafner Publishing Co., New York.

Rao, C, 1965, Linear Statistical Inference and Its Applications, Wiley Springer-Verlag, New York.

Any textbook of Econometrics (the Statistical Appendix).

Dhrymes, P., 1974, Econometrics Statistical Foundations and Applications, Harper & Row Publishers.

Johnston, J., 1980, Econometric Methods, McGraw-Hill.

Judge, G., R. Carter, and associates, 1995, The Theory and Practice of Econometrics, John Wiley & Sons.

Maddala, G., 1990, Econometrics, McGraw-Hill.

LECTURE No IV

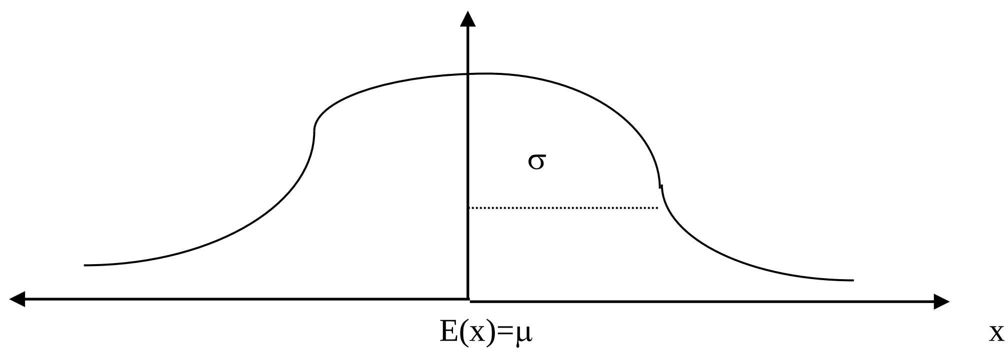
STANDARD DENSITY FUNCTIONS

Thus far, we have discussed some features and estimators of the random variables. This lecture is concerned with the probability/density/distribution functions for continuous random variables and some for discrete random variables.

1. Normal Distribution (x). Let x be an univariate random variable, x is said to be a normal random variable if its distribution is:

$f(x)= [2.\pi.\ \sigma^2\ ]^{-1/2}\ .\exp\{\ (-1/2).(1/\sigma^2).\ [x-E(x)]^2\ \}$; where

$E(x)=\mu$ is the population mean of x; σ^2 is the variance of x; exp=e the neperian number and π is the pi number.



2. Standard Normal Distribution (z). It is a normal distribution with E(z)=0; V(z)=1. The density function is:

$f(z)= [2.\pi]^{-1/2}\ .\exp\{\ (-1/2).z^2\ \}$;

3. Multivariate Normal Distribution (X). Let X={x1, x2,….xn} a set of n random variables where E(X)= Θ and V(X)= Ω. Note that X is a vector of nx1 and V(X) is the matrix of variance and covariance of X of order nxn. It should be noted that:

$V(X)=[vij]= [\ E[(xi-E(xi)).(xj-E(xj)]$;   i, j=1, …n

The density function is:

$f(X)=\ )= [2.\pi]^{-1/n}.\ [/\Omega/]^{-1/2}\ .\exp\{\ (-1/2).[X-E(X)]'.\Omega^{-1}.\ [X-E(X)]\ \}$;

$/\Omega/$= determinant of the Ω.

4. Multivariate Standard Normal Distribution (Z). Let Z={z1, z2, ….zn} the set of n standard normal distribution where E(X)= Θ =0 and V(X)= Ω =In. That is the variables are linear and also independents random variables. The density function is:

$$f(Z)=)= [2.\pi]^{-1/n}. \exp\{ (-1/2).Z'Z \};$$

5. χ^2 (Chi) Square Distribution. Let Z={ z1, …..zn} a set of n independent standard normal distributions then:

$$\chi^2 n= \sum_{i=1}^{n} zi^2;$$ is a Chi Distribution with n degrees of freedom

Its density function is:

$$f(\chi^2 n)= 2^{1/n}.[\Gamma (n/2)]^{-1}. [\exp \{ (1/2). \chi^2 n\}]. [\chi^2 n]^{(n/2-1)} ;$$

Where $\Gamma (n/2)= \int_0^{\infty} u^{(n/2-1)}. \exp\{-u\}.du$; is called the Gamma function.

Some useful properties of this function are:

- As n tend to infinite the Chi square random variable tend to a the standard normal distribution;
- $E(\chi^2 n)=n$;
- Median and Mode of $\chi^2 n= n-2$;
- $V(\chi^2 n)= 2.n$

Kendall (1952), Chapter 12.

6. t-Student Distribution. Let z a standard normal distribution and χ^2 n a chi square random variable with n degree of freedom and independent of z, then the t-Student distribution with n degrees of freedom is defined by:

$tn= z/[\chi^2{}_n / n]^{1/2}$; the density function is given by

$$f(tn)= \Gamma [(n+1)/2]. [n. \pi. \Gamma (n/2)]^{-1/2}. (1+ tn^2/n)^{-(n+1)/2}; \text{ and}$$

Γ (.) is the Gamma function defined above. It is possible to show that: E(tn)=0; V(tn)=n/(n-2); n >2. As n goes to infinite tn goes to z. [Wilks, 1950, Mathematical Statistics, Princeton University Press].

7. The F(d1, d2) Distribution[4]. Let be χ^2_{d1} and χ^2_{d2} two independent chi square distributions with d1 and d2 degrees of freedom respectively then:

$F(d1, d2)= [\chi^2_{d1} /d1]/ [\chi^2_{d2} /d2]$; Its density function is:

$f(F(d1,d2)=\Gamma[(d1+d2)/2]. [\Gamma (d1/2). \Gamma (d2/2)]^{-1}. (d1/d2)^{d1/2}. F^{(d1/2)-1}. (1+ (d1/d2).F)^{-(d1+d2)/2}$

$E(F^r)= \Gamma[(d1/2)+r]. \Gamma [(d2/2)-r]. [\Gamma (d1/2). \Gamma (d2/2)]^{-1}. (d2/d1).r;$

2.r < d2; so for r=1, d2 >2 and E(F)= d2/(d2-2);
Note that $V(F)= E(F^2)- E(F)^2$. So that $V(F)= 2.d2^2. [d1 +d2-2].[d1.(d2-2)^2. (d2-4)]^{-1}$;

[Wilks, 1950, Mathematical Statistics, Princeton University Press and Kmenta, J., 1986, Elements of Econometrics].

8. Distributions from Discrete Random Variables.

8.1 Bernoulli. Let p the probability of occurrence of event represented by x=1 and 1-p the probability of the not occurrence of the same event represented by x=0, then:

$P(x)= p^x .(1-p)^{1-x}$; x ={0,1} is a Bernoulli random variable

E(x)= p; V(x)= p(1-p).
8.2 Binomial. Suppose that we have a sample of n Bernoulli random variables. Where x is the number that the event occurs (or the success event) and n-x is the number that the event does not occur (or the failure event), then P(x) is:

$P(x)= n!/[(n-x)!. x!]. p^x .(1-p)^{n-x}$

Where n!= n.(n-1).(n-2)…2.1

E(x)= n.p; V(x)=n.p.(1-p); some times q=1-p.

[4] Wilks (1950) called the Snedecor's distribution. According to Wilks the Fisher distribution is a transformation of the fomer. If Fi is the Fisher then Fi= (1/2). ln F(d1, d2).

8.3 Poisson. It is a random variable derived from the binomial when n and 1/p tends to infinite and n/p tends to a defined value λ. Usual examples of this is when n is a time period and x the Poisson random variable is the number of occurrences of a particular event within a determined time period.

$P(x)= \exp\{-\lambda\}. \lambda^x / x!; E(x)= V(x)=\lambda.$

REFERENCES

-Any textbook of Statistic.

Anderson, T., 1958, An Introduction to Multivariate Statistical Analysis, Wiley.

Cox, D., D. Hinkley, 1974, Theoretical Statistics, London: Chapman and Hall.

Creighton, J., 1994, A First Course in Probability Models and Statistical Inference.

Kendall, M., 1952, The Advanced Theory of Statistics, Hafner Publishing Co., New York.

Rao, C, 1965, Linear Statistical Inference and Its Applications, Wiley Springer-Verlag, New York.

- Any textbook of Econometrics (the Statistical Appendix).

Dhrymes, P., 1974, Econometrics Statistical Foundations and Applications, Harper & Row Publishers.

Johnston, J., 1980, Econometric Methods, McGraw-Hill.

Judge, G., R. Carter, and associates, 1995, The Theory and Practice of Econometrics, John Wiley & Sons.

Maddala, G., 1990, Econometrics, McGraw-Hill.

LECTURE No V

STATISTICAL INFERENCE AND TEST OF HYPOTHESIS

Thus far, we have revised briefly some of the features and types of random variables and distributions. From now on, our focus is to draw inferences from a set of data or a random sample. We will have two set of variables: Ygxn is the set of g random variables that the theoretical/models/ hypothesis tries to explain (i.e., endogenous variables) and the set of X kxn k random variables that those models use to explain Y. The size of the sample is n. The statistical inferences that we can draw from the set (Y, X) are of the following kinds:

1. Differences on the levels of the Y variables. Assume that Y1 is the wage of a man and Y2 is the wage of woman. Controlling for the different features of these two individuals, we could compare Y1 and Y2 and find a statistical way to find out if they are different. This is done through test of hypothesis that compares two means;

2. Define a statistical and significant interval for a particular variable. The interval are called confidence intervals;

3. To analyze the statistical significance of an explanatory variable. This is done through hypothesis testing on the parameters of a particular function that relates Y and X. For example, $Y=F(X, Y, U, \Lambda)$. Wherein Λ are the set of parameters to be estimated and tested;

4. To estimate a series of parameters of interest. This is done through estimations methods;

5. To find out the time series features of the random variables. This is done through estimation and hypothesis testing. Here we have as examples: causality tests; unit root tests; co integration tests and so on;

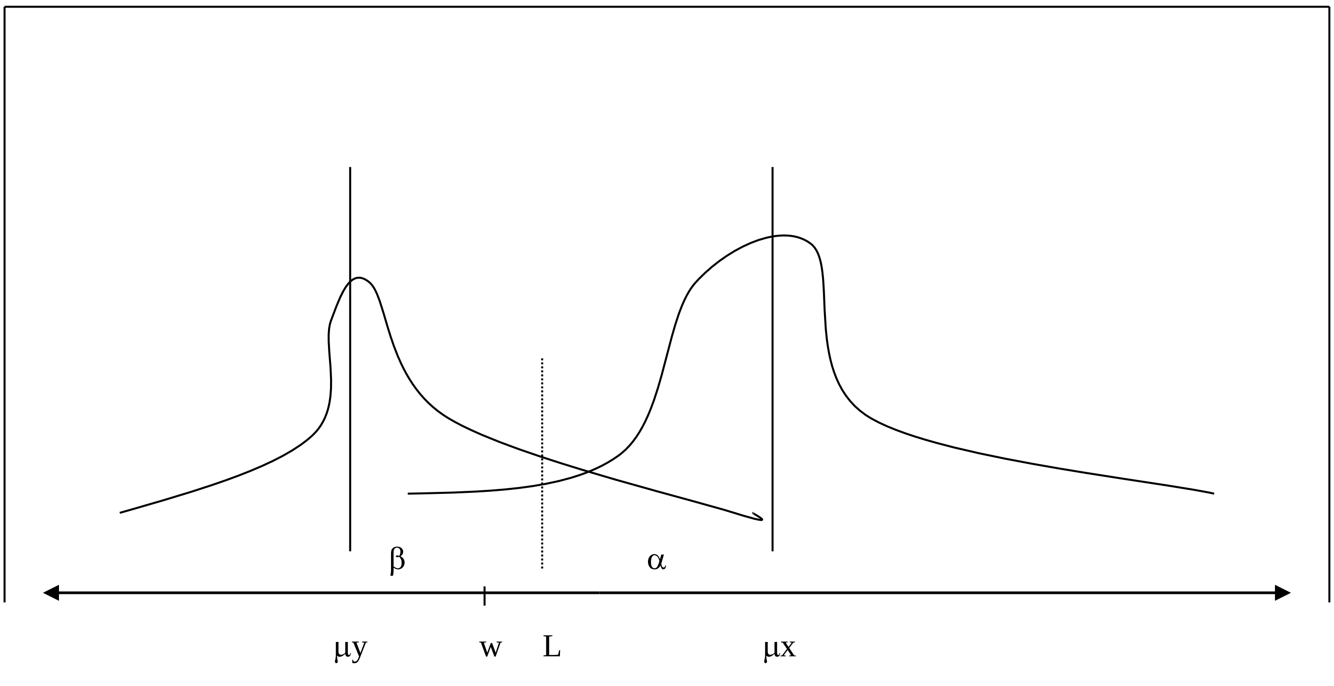
6. To predict Y values using X variables. This is done through estimation and predictors of Y;

1. FUNDAMENTALS OF TESTING HYPHOTESIS:

Suppose that we have two random variables x and y. The symbol ~ means that the distribution of any set of random variables so that $x \sim N(\mu x, \sigma x^2)$ and $y \sim N(\mu y, \sigma y^2)$. N means the normal

distribution. We will assume just for the purpose of this exercise that μx > μy. As we can see from the figure No 3 any value of x or y could belong to any of these two distributions.
Suppose, that w is an estimator of μy and μx ¿ How do we know to which distribution

Figure N° 3:
Hypothesis test



we belong to? Graphically, we can say that the closer is w to any of these two parameters, the higher would be the probability that w belongs to the distribution which expected value is the closer parameter. In Figure No 3, it seems that is closer to μy than to μx. But, ¿how sure could we be? ¿What would be the probability to be wrong? The statistics provides a method to answers these questions. The steps are the following:

1. Define the two types of errors that we could have. Error type I=α= Probability of rejects a hypothesis that is true. Error type II=β= Probability of accepting a hypothesis that is false. In the real world of course, to commit an error type II will have more consequences than committing error type I. In Figure No 3, L is an upper limit number. Whenever that a w estimate that is lower than or equal to L, we will say that w belongs to the variable which expected value is μy. If w is

higher than L we will say that w belongs to the distribution which expected value is μx. Under this L:

α= P(w > L/ E(w)= μy); β= P(w$\leq$ L/E(w)= μx)

"/" = Means the conditional probability assuming the expected value of w be either μy or μx. It should be noted that if L increases, error type II will be higher and error type I will be lower. The contrary happens if L decreases. What this means is that we can not be able to reduce both error simultaneously[5]. From the two types of errors, error type II is the worse, then we may want to minimize this error subject to a fixed low value of error type I (in practice these values are: 10%; 5% and 1%). From this optimization, L will be defined. L will depend upon the two implicit hypotheses. That is that the real distribution is that with E(w)= μy or that where E(w)= μy; the level of α (called significance level) and the sample size, since w is drawn upon a sample and not from the whole universe.

2. Once that L is defined we can make the respective inference.

To complete these two steps we need to define the Hypotheses. One called the null hypothesis that is selected in such as way that is by mistake is rejected the consequences are less harmful than if the alternative hypothesis by mistake is rejected. In this particular, example:

Ho : E(w)= μy= a; [null hypothesis]
H_1 : E(w)= $\mu x \neq a$ [Alternative hypothesis] . This is a 2-tails test;
H_2 : E(w)= $\mu x \geq \mu y$=a [Alternative hypothesis]. This is a right side tail test;
H_3 : E(w)= $\mu x \leq \mu y$ =a [Alternative hypothesis]. This is a left side tail test;

L will depend upon which Hj is selected.

Example I.1 Mean Comparison. Suppose we have a sample of size n of a bivariate population distributed as a N [(μm, μf); $\sigma^{2}\cdot I_2$]. [Note that I_2 is a identity matrix of order 2]. The sample is composed by the incomes of n males and n females. The average of each sample is xm and xf respectively. Both are estimators of the expected values. It can be proved, if σ^2 is known that xm $\sim$ N(μm, σ^2/n) and xf $\sim$ N(μf, σ^2/n). If we want to compare both means then our hypotheses would be:

Ho: μm-μf=0;
H_1: μm-$\mu f \neq 0$;

Our new estimator random variables is d=xm-xm, that can be proved has a ~ N(μm-μf; 2.σ^2/n). Lets define α=0.05, since it is a two tail test the optimum L are actually two values that are symmetric. That is, looking to the standard normal variable L=z0= 1.96 and –L=-zo=-1.96. It can be proved that is the optimal interval that minimizes error type two. If d is a normal random variable it can be easily transformed to a standard normal through:

$z= [d- E(d]/ V(d)^{1/2}$;

$z0=1.96= [d-E(d)] / V(d)^{1/2}$. Under the null hypothesis E(d)=0; V(d)= 2.σ^2/n, then the correspondent upper limit du= $1.96.(2.\sigma^2/n)^{1/2}$. + 0 and the lower limit dl= -1.96. $(2.\sigma^2/n)^{1/2}$. Suppose that the estimate d=0.3, n=100 and σ^2 =25. Then dl= -1.386 and du= 1.386. That is -1.386=dl $\leq$ d $\leq$ 1.386 since d=0.3. The Ho hypothesis is accepted with an error type I of 5%. To compute the error type II we need to define the alternative value for μm-μf . The power of the test is defined as P=1-β. This is the probability of correctly reject the null hypothesis when this in fact is false or when the true distribution is the alternative hypothesis. That is the probability of deciding correctly.

dl and du also define a Confidence Interval (CI) for E(d). We know that:

$P(-zo \leq z \leq zo)=0.95= P(-1.96\leq z \leq 1.96)=0.95$ but

$P(-1.96\leq[d- E(d)]/ 0.707 \leq 1.96)=P(-0.36 \leq 0.3- E(d)\leq 0.36)= P(0.06\leq E(d) \leq 0.66)=.95$

The meaning of the CI is in 95% of the random interval chosen it would contain the population mean.

II. KNOWN DISTRIBUTIONS AND SAMPLE ESTIMATORS:

The key in any hypotheses testing is to find out the relationship between the sample estimators and some known distribution. This section provides some useful properties so that you understand these relationships.

Suppose we have a random sample of a vector X={x1, …x_K} or K random variables. Then if the distribution of each xj is a ~ N(μj, σj^2) and cov(xi, xj)= E[(xi-μi,).(xj-μI)]=0, for i$\neq$j. then:

1. If Y= A.X; where A is a square matrix of order k then:

   E(Y)=A.E(X)=A.Θ; where Θ= {μ1, …. μ_K};

[5] Later on, we will see that both distributions will depend upon the sample size. So in these cases, we can reduce both types of errors by increasing the sample size.

V(Y)= AV(X).A'= A.Ω. A' ; where diag .(Ω)={ $\sigma 1^2$, ….σ_K^2 }; diag. are the elements of the principal diagonal of .Ω; and
Y ~ N[A.Θ; A.Ω.A'];

2. Let xkm be the sample mean of any of the k random variables for a sample of size n then

xkm ~ N[μk ; σk^2 /n];

3. Let sk^2 the sample variance of the xk random variable then:

E(sk^2)= [(n-1)/n]. σk^2 ; that means that sk^2 is a biased estimator of the variance and the unbiased estimator is $s1k^2$ defined above;
V(sk^2) = [(n-1)/n^2]. 2. σk^4
E($s1k^2$)= σk^2 ; that means that $s1k^2$ is a unbiased estimator of the variance ;
V($s1k^2$) = 2. σk^4 /(n-1)

sk^2 ~ (σk^2 /n)χk^2_{n-1} ; or
n. sk^2 / σk^2 = (n-1).$s1k^2$ /σk^2 = ~ χk^2_{n-1} (chi square with n-1 degrees of freedom);

4. With variance unknown then
xkm ~ tn-1 = (xkm- μk)/ s1k. $n^{-1/2}$;

5. The ratio of the sample variances defined as:

(ni/nj). (si^2 / σi^2)/ (sj^2 / σj^2)= (si^2 /sj^2). (σj^2 /(σi^2).(ni/nj);
(si^2 /sj^2). (σj^2 /(σi^2). (ni/nj). [(nj-1)/(ni-1)]=
($s1i^2$ /$s1j^2$). (σj^2 /σi^2)= F(ni-1, nj-1)= [χi^2_{ni-1} /(ni-1)]/ [χj^2_{nj-1} /(nj-1)]

6. Multiple Comparisons and the ANOVA Fundamentals

Suppose now we have two set of random variables K1 for X and K2 for Y={y1, y2, ..y_{K1}}. We can use a table for crossing these variables:

	x1	x2	xi	x_{K1}
y1				y1.
y2				
..				
y_{K2}				$y_{K2.}$
	x.1		$x_{.K1}$	[x..= Σ $x_{.j}$] ; [y..=Σ yi .]

where yi. is sum of the random variable yi across all the random variables x;
x.j is the sum of the random variable xj across all the random variables y; x.. or y.. is the sum of all the values x and y.
Let xm, and ym the average of each set of random variable where the sample size is nx and ny. Then the total sample variance for x is

$$\sum_{i=1}^{K2}\sum_{j=1}^{K1}(xij - xm)^2/(nx-1) \sim (\sigma x^2/(nx-1))\ \chi^2_{nx-1} \text{ or}$$

$$\sum_{=1}^{K2}\sum_{j=1}^{K1}(xij - xm)^2/\sigma x^2 \sim \chi^2_{nx-1}$$

Here it is assumed that V(xij) (= σx^2) is the same for any xij. By the same token, we have:

$$\sum_{i=1}^{K2}\sum_{j=1}^{K1}(xij - xm.j)^2/\sigma x^2 \sim \chi^2_{nx-K1}$$; where xm.j is the average of each

column of x;

$$\sum_{j=1}^{K1}(x.j - xm)^2/\sigma x^2 \sim \chi^2_{K1-1} ;$$

III. ANALYSIS OF VARIANCES, ANOVA

Suppose that a random variable can be divided in K1 groups according to certain features. The question is if these features makes difference for the random variable. One way ANOVA, is the technique that deals with this question because we are testing only one set of K1 features. If there are two sets of features then we have Two-Ways ANOVA and if we have N sets of features we will have a N-ways ANOVA. The principle of the ANOVA is simple. By the former section we know that:

$$\sum_{i=1}^{K2}\sum_{j=1}^{K1}(xij - xm)^2 = \sum_{i=1}^{K2}\sum_{j=1}^{K1}(xij - xm.j)^2 + \sum_{j=1}^{K1}(x.j - xm)^2$$

nx –1 nx-K1 K1-1 [degrees of freedom]

Total (sample Variance)= Sample Variance between Groups Sample variance within each group

The ratio $F(nx-K1, K1-1) = \sum_{i=1}^{K2}\sum_{j=1}^{K1}(xij - xm.j)^2/(nx-K1)]/ \sum_{j=1}^{K1}(x.j - xm)^2/(K1-1)$

$=F \sim F(nx-K1, K1-1) \sim Fi(nx-K1, K1-1)$

Under Ho there is no difference between group son both variances should be same as the ratios should be equal to 1. If very far then there are significant differences between groups. The same principle it is applied for N- ways ANOVA.

IV. MAXIMUM LIKELIHOOD FUNCTION/ESTIMATORS AND HYPOTHESES TESTING

1.The Maximum Likelihood Function. Let X a matrix of K random variables.
draw upon a random sample of size n.That is:
$X= \{x1. ,,,, x_K\}$; further $xjnx1 \sim Df(\Theta j; \Omega j)$. That is Df is the density function of the vector xj, where $E(xj)= \Theta j$ and $V(xj)= \Omega j$; this latter is square matrix of order n. The joint density of the sample for the K random variables is:

$$L = \prod_{j=1}^{n} Dfj$$

The L function, which the joint density of the sample is called the Likelihood Function of the sample. A method of estimation is to find out the estimator of [Θj ; Ωj] such that this function be maximized. The set of estimators are called the MLE estimators. In practice, what it is maximized is not L but rather lnL:

$$\text{Max } \ln L = \sum_{i=1}^{n} \ln DFj;$$

Note that dlnL/d Λ=0 and d^2 lnL/d $\Lambda d\Lambda'$; are also random variables, and $\Lambda=[[\Theta j ; \Omega j]$, j=1,..n

2. Statistics Tests Based on MLE. The following tests come from the estimator Λmlej of ML without any restrictions. Suppose we want to test the following function:

Ho: $h(\Lambda j)=0$; where h is a vector of size $q < K+ n^2$; for j=1, ..n.

Lets assume that $\Lambda j= r(\alpha j)$ under Ho; where αj is a vector of size= $K+ n^2 – q$; so that

Λmleoj is the constrained MLE of Λj under Ho. Then

Wald test:

$W=- h(\Lambda mlej)'. \{ \delta h/\delta\Lambda j'(\Lambda mlej). [\delta^2 \ln L/\delta\Lambda j\delta\Lambda j'(\Lambda mlej)]^{-1} .\delta h'/\delta\Lambda j(\Lambda mlej)\}.h(\Lambda mlej);$

Lilkelihood Ratio Test:

$LRT= -2. Ln[(Max L(\Lambda j)/h(\Lambda j)=0)/(Max L(\Lambda j))]= 2.[\ln L(\Lambda mlej)-\ln L(\Lambda mleoj)];$

Lagrangian Multiplier/Rao/Score Test:

$LM= \delta h/\delta\Lambda j'(\Lambda mleoj).[\delta^2 \ln L/\delta\Lambda j\delta\Lambda j'(\Lambda mleoj)]^{-1}\delta h'/\delta\Lambda j(\Lambda mleoj);$

When the size of the sample goes to infinity these three statistics tend to χ^2 q (degrees of freedom]. [Amemiya, T., 1985, Advanced Econometrics, pp. 141-146].

REFERENCES

-Any textbook of Statistic.

Anderson, T., 1958, An Introduction to Multivariate Statistical Analysis, Wiley.

Cox, D., D. Hinkley, 1974, Theoretical Statistics, London: Chapman and Hall.

Creighton, J., 1994, A First Course in Probability Models and Statistical Inference.

Kendall, M., 1952, The Advanced Theory of Statistics, Hafner Publishing Co., New York.

Rao, C, 1965, Linear Statistical Inference and Its Applications, Wiley Springer-Verlag, New York].

- Any textbook of Econometrics (the Statistical Appendix).

Dhrymes, P., 1974, Econometrics Statistical Foundations and Applications, Harper & Row Publishers.

Engle, R., D. MacFadden, 1994, Handbook of Econometrics, Vol. IV, Chapter 36, “Large Sample Estimation and Hypothesis Testing”. North Holland, Elsevier.

Intrilligator M., Z. Griliches, 1983, Handbook of Econometrics, Vol, II, Part 4, Testing.

Johnston, J., 1980, Econometric Methods, McGraw-Hill.

Judge, G., R. Carter, and associates, 1995, The Theory and Practice of Econometrics, John Wiley & Sons.

Maddala, G., 1990, Econometrics, McGraw-Hill.

LECTURE No VI

ECONOMETRIC METHODS

1.THE CLASSICAL MULTIVARIATE LINEAR REGRESSION MODEL.

Thus far, it has been presented the statistical basis for the econometric methods. The methodology was descriptive presenting results, without a real world interpretation and without proof. From this lecture and on, the methodology will be more intuitive than formal. Though, some formal analysis when necessary will be presented or indicated.

Social science theory deals with the explanation of a determined human event. For example:

1. The factors (Xit) that determines/explain economic growth (Yit);

2. The factors (Xit) that explain the level of activity participation of a particular type of population (females, immigrants, etc., Yit);

3. The factors (Xit) that determines the demand or not for a particular services (e.g., broadband services, inter net services, etc. Yit);

In each case there exist or set of variables Yit that is explained by factors Xit. In econometrics, Yit is the set of endogenous (also called dependent) variables and Xit is the set of exogenous or predetermined (also called independent) variables. In order that econometrics technique is able to draw inferences from the data, it needs a series of assumptions. The inferences then are subject to the validity of these assumptions and the theory used. The simplest model in econometrics, is the objective of this lecture. This is called the Classical Multivariate Linear Regression Model.

1.1 Assumptions:

A1: $Y = X.\beta + \varepsilon$; A Linear relation among parameters of interest

Y is vector of sample of size N of a endogenous variable, i.e., $Y= \{y1, \ldots\ldots, y_N\}$ and X is a matrix of order NxK of K exogenous variables (K < N) from a sample of size N, i.e., $X= [x1, \ldots\ldots., x_K]$, where $xk=\{ x1k, \ldots..x_Nk\}$ is a vector or N random variables; β is vector of K parameters. A1 implies that causality runs from X to Y; ε is a vector of N random variables, i.e., $\varepsilon=\{\varepsilon1, \ldots.\varepsilon_N\}$.

There are two ways to interpret A1. The first is that Y is random variable due to fact that it depends upon the stochastic term ε. In this interpretation X is not a set of random variables. The second is that Y is linear dependent of X but in the conditional probability sense. That is Y, X are both of sets of random variables but that the conditional relationship between Y an X is conditional. In both interpretations the error term ε is of the same nature than Y. We will use this second interpretation for the rest of the lecture;

A2: $E(Y/X)= X.\beta \Leftrightarrow E(\varepsilon/X)=0$; non-specification error.

The symbol "/" means the conditional mean. This assumption means, that the conditional population mean of Y is a linear combination of X and β or that it is expected that the error term does not include any other factor relevant to explain Y;

A2.1: $E(X'\varepsilon/X)=0 \Leftrightarrow E(\varepsilon/X)=0$; non-simultaneous equation error. This assumption means that the conditional covariance between X and ε is zero. That is, there is no linear relationship between X and the error term. Under, the conditional interpretation, both assumptions are equivalents;

A3: rank(X)=K; Non-multicolinearity among exogenous variables.

This means that the K random variables are linearly independent. X is also called the designed Matrix;

A4: $V(Y/X)= E(\varepsilon.\varepsilon'/X)=\sigma^2. In$. Homocedastic and Non-Autocorrelated error.

This assumption means that the variance of any particular observation $V(yi/X)$ or the variance of the error term is constant and that there is no linear dependence between any pair of observation (Yi, Yj) or error. That means $Cov(Yi,Yj/X)= Cov(\varepsilon i, \varepsilon j/X)=0$; In is the identity square matrix of order n;

A5: $Y/X \sim N(X.\beta; \sigma^2. In.) \Leftrightarrow \varepsilon/ X \sim N(0; \sigma^2. In.)$.

The conditional random variable Y (or ε) is normal distributed.

1.2 The Minimum Least Squares Estimators (MLS or LS):

To make inferences from the data we need to know or estimate β and σ^2. So we need to find suitable estimators for both set of parameters. Estimators are infinite but estimators that have sound statistical property are finite. Those statistical properties exist for both finite and infinite sample sizes. The latter are also called asymptotic properties of the estimators. We will introduce these properties, as we need it. The most known estimators for the classical model is the MLS. This method comes from:

$$\min_{\beta} \varepsilon'\varepsilon = (Y-X.\beta)'(Y-X.\beta) = \sum_{i=1}^{N} \varepsilon i^2 = \sum_{i=1}^{N} (Yi- Xi.\beta)^2 ;$$

$d(\varepsilon'\varepsilon)/d\beta = -Y'X - Y'X + 2.(X'.X).\beta=0 \Leftrightarrow (X'.X).\beta e= X'.Y$ [Normal equations]; [1]
$d^2((\varepsilon'\varepsilon)/d\beta.d\beta' = (X'X)$. This square matrix of order k (under A3) is definite positive.

From [1] we have:

[2] $\beta e= (X'.X)^{-1}. X'Y$; LS of β

Replacing A1 in [2] we have

$\beta e= (X'.X)^{-1}. X'.[X.\beta+ \varepsilon] = \beta+ (X'.X)^{-1}. X'.\varepsilon$;

$E(\beta e/X)= \beta$; that means the LS estimator is an UNBIASED estimator of β;

$(\beta e -\beta)= (X'.X)^{-1}. X'.\varepsilon$; then

$V(\beta e)= E[(\beta e -\beta).(\beta e -\beta)']= \sigma^2. (X'X)^{-1}$;

$\beta e \sim N[\beta; \sigma^2. (X'X)^{-1}]$; since the LS is a linear function of ε ;

The LS of the error term is:

$\varepsilon e = Y-X.(X'.X)^{-1}.X'.Y= [In- X. (X'.X)^{-1}.X'].(X.\beta +\varepsilon)$
$= [In- X. (X'.X)^{-1}.X'].\varepsilon) =M. \varepsilon$; $M=In- X. (X'.X)^{-1}.X'$; this is an idempotent and symmetric matrix [i.e., since AA'=A'A=A]. For any idempotent matrix the rank of this matrix is equal to the trace (tr) of the matrix. Then:

$tr(M)=tr(In)- tr(X. (X'.X)^{-1}.X')= N- tr(I_K)=N-K$; since tr(A.B)=tr(BA); as ;long as B and A can be multiplied. This implies that εe is a linear combination of N-K linear independent ε vectors. This implies that:

$$\varepsilon e'\varepsilon e =\sum_{i=1}^{N} \varepsilon ei^2 = \varepsilon'M \varepsilon$$ is a sum of N-K normal distributions;

$E(\varepsilon e'\varepsilon e)= E[tr(\varepsilon' M\ \varepsilon)]= \sigma^2. (N-K)$; so an unbiased estimators of σe^2 is:

$\sigma e^2 = \varepsilon e'\varepsilon e/(N-K)$: since $E(\sigma e^2)= \sigma^2$

$V(\varepsilon e)= E(\varepsilon e.\varepsilon e')= \sigma^2. M$; $\varepsilon e\sim N(0;)= \sigma^2. M$

It can be proved that:

$\varepsilon'.M.\varepsilon/\sigma^2 = \varepsilon e'.\varepsilon e/\sigma^2 \sim \chi^2$ with N-K degrees of freedom. Using this result and the fact that βe is a normal distribution then:

$(\beta ei- \beta i)/[\sigma e^2\ (x'ixi)^{-1}] \sim t_{N-K}$; where $x'ixi$ is the ii element of the main diagonal of the matrix $(X'X)^{-1}$.

Gauss-Markov Theorem. Under A1-A5, the LS of β is the BLUE [i.e, the Best Linear Unbiased, Estimator]. Above it has been proved that βe is linear on ε or Y and it is unbiased. What is remaining to be proved is that is the LS estimator within the class of all linear unbiased estimators is the one that has the minimum variance.

Suppose we have another estimator of β, b so that:

$b= A.Y$; such that $E(b)= A.X.\beta=\beta$; $AX=I_K$; then:
$V(b)= AV(Y/X).A= \sigma^2. A.A'$;

Let $D= (X'.X)^{-1}. X'- A$; then $DX= 0$ and $(X'.X)^{-1}. X' =D+A$; then

$V(b)= \sigma^2. [-D+ (X'.X)^{-1}. X'].[-D' + X.(X'.X)^{-1}]= \sigma^2. [D.D' + (X'.X)^{-1}]$

$V(b)= \sigma^2. (D.D') + V(\beta e)$; Note that $D.D'$ is a semi definite positive matrix; so that:

$V(b)- V(\beta e) \geq 0$; so that the LS of βe has the minimum variance within its class.

1.3. Hypotheses Testing:

Suppose that we want to test the following general hypothesis

Ho: $R\beta= \theta$; R is a matrix of order rxK; and θ is a vector of order rx1.
H1: Ho not true;

We know that :

$R\beta e \sim N\,[\theta\ ;\ \sigma^2.\ R.\ (X'.X)^{-1}.R']$; since $R\beta e$ is a linear function of a normal random variables. Then the sum of r standard normal random variables are:

$z'z= (R\beta e- \theta\ )'\ .[V(R.\beta e)]^{-1}.\ (R\beta e- \theta\ ) \sim\chi^2$ r degrees of freedom and;

$\varepsilon e'.\varepsilon e/\sigma^2\ \sim\ \ \chi^2\ (N\text{-}K)$ degrees of freedom (see above). The

$F_{r,\,N-K} =\ \ [\ \chi^2\ r\ /r]/\ [\chi^2{}_{N-K}/\ (N\text{-}K)]$

$F_{r,\,N-K} =\ \ [(R\beta e- \theta\ )'\ .[V(R.\beta e)]^{-1}.\ (R\beta e- \theta\ )/r]/\ [(\varepsilon e'.\varepsilon e/\sigma^2)/(N\text{-}K)]$

$F_{r,\,N-K} =\ \ [(R\beta e- \theta\ )'\ .(1\ /\sigma^2).\ [R.\ (X'.X)^{-1}.R']^{-1}.\ (R\beta e- \theta\ )/r]/\ [(\varepsilon e'.\varepsilon e/\sigma^2)/(N\text{-}K)]$

$F_{r,\,N-K} =\ \ [(R\beta e- \theta\ )'.\ [R.\ (X'.X)^{-1}.R']^{-1}.\ (R\beta e- \theta\ )/r]/\ \sigma e^2$

Let βe_R= be the LS of the restricted model (i.e., when Ho holds) and $\beta e=\beta e_u$ the LS of the unrestricted model (when H1 holds). In can be proved that LS subject to Ho is

$\beta e_R= \beta e - (X'X)^{-1}.\ R'.[R.\ (X'X)^{-1}.R']^{-1}.\ (R.\ \beta e\ -\ \theta)$; and

$F_{r,\,N-K} = [(N\text{-}K)/r].\ \ [\ (\varepsilon e_R'\ \varepsilon e_R\ -\ \ \varepsilon e_u'\ \varepsilon e_u\ )/\ (\varepsilon e_u'\ \varepsilon e_u\ )]$

Example 1.3.1. Individual coefficients

R= [rij], rii=1; rij=0 for $i\neq j$; i=1, r;j=1,…..K; then:

Ho: $\beta j=0$;
H1: Not Ho;

r=1; and $t_{N-K}{}^2 = F_{1,\,N-K}$; under Ho; $t_{N-K} = \beta ej/\ [V(\beta ej)]^{1/2}$;

Example 1.3.2. Total set of coefficients.

Ho:β=0; $R=I_K$
H1: Not Ho;

$F_{K,\,N-K}\ /Ho = (\beta e'.(X'.X).\beta e)/K]/\ \sigma e^2 = [Ye'Ye/K]/\ \sigma e^2$;

Where $Ye'Ye = \Sigma\, Ye^2$; the Total of Sum of the Estimators of Y= Total Square Sum of the regression

1.4 Determination Coefficient.

a. Some Useful Identities:

Y'Y= Total Sum Squares; $\varepsilon e'\varepsilon e$= Total Sum of Square Errors

$Y = Ye + \varepsilon e$; $Y'Y = Ye'Ye + \varepsilon e'\varepsilon e$;

Let $Md = I_N - i.i'/N$; $i = \{1, \ldots\ldots 1)$ a vector of ones of order Nx1; trace (ii')=N;
Md is a idempotent matrix of order NxN
And trace(Md)= rank (Md)=N-1;

$Y = X\beta + \varepsilon$; pre multiplying by Md

$Md.Y = Md.X\beta + Md.\varepsilon$;

$MdY = \{Y_1 - Ym, \ldots.. Y_N - Ym) = Yd$; and $Ym = \Sigma\, Yj/N$ the sample mean of Y;

Yd are the set of data in deviations so that the new model is:

$Yd = Xd.\beta d + .\varepsilon d$: Note if there is an intercept then $x1j = i = \{1\ldots1)$ and $x1m=1$ so actually we will have only K-1 parameters since we are eliminating the intercept parameter, so that $\beta d = [\beta 2, \ldots\ldots\beta_K]$:

$\beta 1e = Ym - Xm.\beta d$; where Ym and Xm are the set of the sample means of the variables.

Note also that because of the normal equations $\varepsilon em=0$; so that $\varepsilon e'\varepsilon e = \varepsilon de'\varepsilon de$;

En terms of data en deviations we have that

$Yd'Yd = Yde'Yde + \varepsilon e'\varepsilon e$;

b. Some Properties of the Normal distribution:

Lets assume that the pair set of random variables $(Y, X) \sim N[(\Theta y, \Theta x); \Omega]$; where

$$\Omega = \begin{pmatrix} \Sigma y & \Sigma yx \\ \Sigma xy & \Sigma x \end{pmatrix}$$; the matrix of variances and covariances between Y and X;

Further $V(Y) = \Sigma y$; $V(X) = \Sigma x$; $Cov(Y,X) = Cov(X,Y) = \Sigma xy$

Then $Y/X \sim N[(\Theta y + \Sigma yx. \Sigma x^{-1}. (X - \Theta x); \Sigma y - \Sigma yx. \Sigma x^{-1}. \Sigma xy]$

The population multiple correlation or determination coefficient is the maximum simple correlation coefficient between Y and the linear combination $X_{1xK}. \beta_{Kx1}$; so that β is the vector that maximize that correlation.

Note that $E'E = (Y - X\beta)'(Y - X\beta)$ and

$V(E'E) = V(Y'Y - Y'X.\beta - \beta'X'Y + \beta'X'X\beta)$

The parameter β that minimize this variance is $\beta = \Sigma x^{-1}. \Sigma xy$

Then population multiple correlation coefficient is defined by:

$R^2 = [\beta'. \Sigma xy]/[\Sigma y] = [\Sigma xy' \Sigma x^{-1}. \Sigma xy]/\Sigma y$

The sample multiple correlation coefficient is defined by:

$Re^2 = [\beta e'.X'.Y]/(Y'.Y) = [\beta de'.Xd'.Yd]/(Yd'.Yd) = [Y'.X. (X'.X)^{-1}X'.Y]/(Y'.Y) =$

$= Yd'.Xd\,(Xd'.Xd)^{-1}Xd'.Yd]/(Yd'.Yd) = (Ye'.Ye)/(Y'.Y) = (Yde'.Yde)/Yd'.Yd$

Note that: $0 \leq Re^2 \leq 1$. Re^2 is also used as a measure of fit of the regression model. The closer Re^2 to one the degree of fit is higher. This does not means that the model is working better as Re^2 is close to one. Since, sometimes we could have spurious regression, as we will se below.

1.5 The Maximum Likelihood Estimator (MLE)

As we have seen from chapter V. MLE is another well-known good estimator used often in econometrics usually in models of large samples. We will apply its definition for the classical model that we are analyzing in this chapter.

The Likelihood function for the sample Y/X is:

$$L(Y/X, \beta, \sigma^2)= \prod_{i=1}^{N} [2.\pi.\ \sigma^2\ ]^{-1/2}\ .\exp\{\ (-1/2).(1/\sigma^2).\ [Yi-Xi.\ \beta]'[Yi-Xi.\ \beta]\ \};\ \text{where}$$

$$L= [2.\pi.\ \sigma^2\ ]^{-N/2}\ .\exp\{\ (-1/2).(1/\sigma^2).\ [Y-X.\ \beta]'[Y-X.\ \beta]\ \};\ \text{where}$$

$$\ln L= -\ \{N/2.\ \ln(2.\pi.\ \sigma^2\ ) + (1/2).(1/\sigma^2).\ [Y-X.\ \beta]'[Y-X.\ \beta]\}$$

Note that Max lnL is equivalent to minimize the sum of square errors so that the LS and the MLE yield the same estimators for the classical multi regression model. However, the estimator of the variance is different. First, lets take the first derivative with respect to (β, σ^2):

[1] $d\ln L/d\beta= (1/2).(1/\sigma^2).(2).\ [Y'X - (X'.\ X).\beta]=0$

[2] $d\ln L/d\sigma^2 = -(N/2).\ [2.\pi.\ \sigma^2\ ]^{-1}\ .(2.\pi) + (1/2).(1/\sigma^2)^2.\ [Y-X.\ \beta]'[Y-X.\ \beta]=0$

[3] $d^2\ln L/d\beta\ d\beta'= -(1/\sigma^2).(X'X);$

[4] $d^2\ln L/d(\sigma^2)^2;= (N/2).\ (\sigma^2\ )^{-2} - (\sigma^2)^{-3}.\ [Y-X.\ \beta]'[Y-X.\ \beta]=;$

[5] $d^2\ln L/d\beta d\sigma^2 = 0= [d^2\ln L/d\sigma^2\ d\beta]';$

From [1] we have that:

$\beta e=\beta_{MLE}= (X'X)^{-1}.\ X'Y;$

From [2] we have:

$\sigma^2_{MLE} = (\varepsilon e'\ \varepsilon e)/N$; so it is a biased estimators of σ^2;

$$V_{MLE}(\beta, \sigma^2) = (-1)\ .\begin{pmatrix} d^2\ln L/d\beta\ d\beta' & d^2\ln L/d\beta d\sigma^2 \\ d^2\ln L/d\sigma^2\ d\beta & d^2\ln L/d\beta d\sigma^2 \end{pmatrix}^{-1}$$

; the matrix of variance and covariance of MLE (β, σ^2).

$$V_{MLE}(\beta, \sigma^2)= \begin{pmatrix} \sigma^2.(X'X)^{-1} & 0 \\ & \\ 0 & (2/N).\ \sigma^4 \end{pmatrix}$$

Some important asymptotic properties of the MLE are:

Under the regularity conditions which are:

i) The range of the random variables (Y, X) are independent of the parameters Λ;

ii) The f function of the sample possesses derivatives of at least third order with respect to Λ, and these derivatives are bounded by integrable functions of (Y, X). This means that:

$/ d^3L/d\theta^3 / \leq$ M(Y,X) and E(M(Y,X) <P for some constant, θ is a parameter of Λ.

1. *$V(dlnL/d\Lambda)= - E\ [d^2lnL/d\Lambda d\Lambda']$;* is the matrix of variance and covariance of the first derivative of the lnL with respect to Λ; and *$E[dlnL/d\Lambda]=0$;*

2. At least one root of $dlnL/d\Lambda=0$ is a consistent estimator. That is:

$\text{Plim}_{N\to\infty}\ \theta_{MLE} = \theta; \Leftrightarrow \text{Lim}_{N\to\infty}\ \text{Prob}(\ /\ \theta_{MLE} - \theta / < \delta) > 1-\kappa$; for small values of δ and κ

3. Let be $z= N^{1/2} .[\ \theta_{MLE} -\theta]$ then $z \sim A\ N(0, - [E(N^{-1}.\ d^2lnL/d(\theta^2)^2)]^{-1}$;

$\sim$A means asymptotically when $N\to\infty$;

4. The MLE is the Minimum Variance Bound estimator of the class of Consistent, Unbiased and Asymptotically Normal (CUAN) distributed estimators. In the sense, that it reaches the Cramer-Rao matrix of variance and covariance and equal to:

$$- [E(N^{-1}.\ d^2lnL/d(\theta^2)^2)]^{-1} = - [N.M(\theta)]^{-1}.$$

2 and 3 it means that the MLE is consistent and efficient since it has a minimum variance.

1.6 W, LM, LR in the Classical Multi regression Model.

Let Ho: $R.\beta=0$ the null hypothesis then:

$h(\Lambda mlej)= R.\beta_{MLE}$; $dh/\delta\Lambda j'(\Lambda mlej)= d(R.\beta)/d\beta= R$;

$[\delta^2 lnL/\delta\Lambda j\delta\Lambda j'(\Lambda mlej)]^{-1} = [d^2 lnL/d\beta\ d\beta']^{-1} = \sigma_{MLE}^2 .(X'X)^{-1}$

We know that the Wald test is defined by:

$W=- h(\Lambda mlej)'. \{ \delta h/\delta\Lambda j'(\Lambda mlej). [\delta^2 lnL/\delta\Lambda j\delta\Lambda j'(\Lambda mlej)]^{-1} .\delta h'/\delta\Lambda j(\Lambda mlej)\}.h(\Lambda mlej);$

$W= - \beta_{MLE}'.R'.\{ R.[\sigma_{MLE}^2 .(X'X)^{-1}].R'\}.R. \beta_{MLE}' \sim\chi^2 r$

We now that under the null hypotheses that $\theta=0$

$$F_{r, N-K} = [(R\beta e)'. [R. (X'.X)^{-1}.R']^{-1}. (R\beta e)/r]/ \sigma e^2$$

$$= \beta e'R' [R. \sigma e^2 (X'.X)^{-1}.R']^{-1}. (R.\beta e)/r]$$

We know that $\beta e= \beta_{MLE}$ and $\sigma e^2= N/(N-K). \sigma_{MLE}{}^2$ replacing we have:

$F_{r, N-K} = (N-K)/(N.r).W$; $F \leq W$, note when $N\rightarrow\infty$ then $F=W/r$

Using the same hypothesis we have that:

$lnL (\beta_{MLE} , \sigma_{MLE}{}^2)= lnL= - \{N/2. ln(2.\pi. \sigma_{MLE}^2) + (1/2).(1/\sigma_{MLE}^2). [Y-X. \beta_{MLE}]'[Y-X. \beta_{MLE}]\}$

$lnL= - \{N/2. ln(2.\pi. \sigma_{MLE}^2) + (1/2).(1/\sigma_{MLE}^2). \varepsilon_{MLE}'.\varepsilon_{MLE}\}=lnLu$ the unrestricted model

lnL Under Ho=lnL (Ho)=lnL_R the restricted model

But the LRT is defined by:

$LRT= -2. Ln[(Max\ L(\Lambda j)/h(\Lambda j)=0)/(Max\ L(\Lambda j))]= 2.[lnL(\Lambda mlej)-lnL(\Lambda mleoj)];$

So applying we have:

$LRT= 2.[lnLu-lnLr]=2ln[Lu/Lr]= 2. Ln \{[\sigma_{MLE\ R}{}^2/\sigma_{MLE\ U}{}^2]^{-N/2}. Exp\{ (1/2).(1/\sigma_{MLER}^2). \varepsilon_{MLER}'.\varepsilon_{MLER} - (1/2).(1/\sigma_{MLEU}^2). \varepsilon_{MLEU}'.\varepsilon_{MLEU}\}$

$LRT= 2.[(N/2).ln[\sigma_{MLE\ R}{}^2/\sigma_{MLE\ U}{}^2] + (1/2).(1/\sigma_{MLER}^2). \varepsilon_{MLER}'.\varepsilon_{MLER} - (1/2).(1/\sigma_{MLEU}^2). \varepsilon_{MLEU}'.\varepsilon_{MLEU}];$

But $\sigma_{MLEU}^2 = \varepsilon_{MLEU}'.\varepsilon_{MLEU}/N$; the same for the restricted model, then:

$LRT = N.\ln[\sigma_{MLE\ R}{}^2/\sigma_{MLE\ U}{}^2] = N.\ln[\ \sigma e_R{}^2/\sigma e_U{}^2] = \ \sim \chi^2 r$

Applying the same for the LM test we have:

$LM = \delta h/\delta\Lambda j'(\Lambda mleoj).[\delta^2 \ln L/\delta\Lambda j\delta\Lambda j'(\Lambda mleoj)]^{-1}\delta h'/\delta\Lambda j(\Lambda mleoj);$

It can be proved that:

$LM = N.[\ \varepsilon e_R'.\varepsilon e_R - \varepsilon e_u'.\varepsilon e_u]/\ \varepsilon e_R'.\varepsilon e_R\ ;$

$W = N.[\ \varepsilon e_R'.\varepsilon e_R - \varepsilon e_u'.\varepsilon e_u]/\ \varepsilon e_u'.\varepsilon e_u\ ;$

$LRT = N.\ln\{\ \varepsilon e_R'.\varepsilon e_R/\ \varepsilon e_u'.\varepsilon e_u\};$

$F_{r,\ N-K} = [(N-K)/r].\ \ [\ (\varepsilon e_R'\ \varepsilon e_R - \ \varepsilon e_u'\ \varepsilon e_u\ )/\ (\varepsilon e_u'\ \varepsilon e_u\ )]$

$W \geq LRT \geq LM \geq F$ {Unde Ho: $R\beta=0$}

W,LRT, LM ~ $\chi^2 r$

[Amemiya, T., 1984, Advanced Econometrics, pp. 144]

1.7 Final Remarks

The Classical Model is useful for the following purposes:

1. To determine what Xj factors (given by the theory) are important to determine the Y variable. This is the conclusion if the following results are provided:

i) The signs of the coefficient are as predicted by the theory;
ii) The coefficient of Xj is statistically significant; [i.e. , the Ho: $\beta j=0$ is rejected]
iii) the elasticity of Y with respect to Xj is not close to zero

2. To determine what set of X factors are the most important to determine Y. This conclusion is provided whenever 1 is conclude and the test Ho: $\beta s=0$ is rejected for a set of $s \leq K$ explanatory variables;

3. Higher level of Re^2 only means that the fit to a linear relationship is closer. It does not mean that the model is better. Thus, you may have Re^2 closer to 1 with useless models because they are spurious models (as it defined below);

4. The model does not prove causality, on the contrary, the model assume a causality relationship going from X to Y;

5. Conclusions 1 to 4 are valid only if all the assumptions of the classical model holds. If some of those does not hold, then other methods should be applied or it is necessary to correct the model to one where the assumptions holds. These methods are part of the econometric methodology that unfortunately we will not revise here.

6. In practice you do not need to revise tables to determine if Ho is accepted or reject most econometrics software provide the called Prob.-Value. This is the minimum probability of error type I given the values of the estimates. Usually we will reject Ho for Prob.-Value lower than 10%, if you want to be more demanding about your data set or model you could ask for prob.-values lower than 5% to reject Ho.

7. The LS and MLE in the linear model, under the A1-A5 is an unbiased, efficient (i.e, has a minimum variance relative to other estimators within its class), and consistent estimator. Further, it can be proved that the MLE has an asymptotic normal distribution and it is the most asymptotically efficient within its class. If A1 fails to hold, then we need other LS techniques called non-linear least squares, but still we can use the MLE. If A2 does not hold then the LS, MLS are biased and inconsistent estimators and the error is called misspecification error. If A3 does not hold the properties of the LS and MLS do hold, the problem, however, is identification. That is, no longer may exist the LS or MLS estimators. If A4 does not hold then the LS and MLS would be inefficient. Generalized Least Squares would be the efficient estimators. This method means transforming the original data to another set such that the A4 holds. The difference in the statistical methods is on the way that the transformation and estimation takes places. If A5 does not hold no test are valid for a finite small sample. Although for a large sample, still the asymptotic properties and test are valid.

REFERENCES

Dhrymes, P., 1974, Econometrics Statistical Foundations and Applications, Harper & Row Publishers.

Engle, R., D. MacFadden, 1994, Handbook of Econometrics, Vol. IV, Chapter 36, “Large Sample Estimation and Hypothesis Testing”. North Holland, Elsevier.

Intrilligator M., Z. Griliches, 1983, Handbook of Econometrics, Vol I, chapter 2 [“Statistical Theory and Econometrics”, A. Zellner] and 3 [“Economic and Econometric Models”, M. Intrilligator].

Johnston, J., 1980, Econometric Methods, McGraw-Hill;
Judge, G., R. Carter, and associates, 1995, The Theory and Practice of Econometrics, John Wiley & Sons.

Maddala, G., 1990, Econometrics, McGraw-Hill.

2.LIMITED-DEPENDENT AND QUALITATIVE VARIABLES: THE LOGIT AND PROBIT MODELS

Real data do no need to be a continuos variable. It could be integer number and even qualitative data. This section we will deal with these latter types of data. It will be assumed that the theory provide an explanation or a mechanics by which the specification model allows for the use of this kind of data, in particular the endogenous or dependent variables. Usually kinds of qualitative dependent variable:

> **i) Truncated Dependent variables.** When Y takes two set of values Y > yo or Y<yo. Example of this variable is earnings or wages where yo is the minimum earnings;
>
> **ii) Censored Dependent Variable.** When there exist two set of values for Y. Y=0 or Y ≠ 0 (i.e., Y>0 or Y<0). For example, households ‘ expenditures for some good. There are some households that will not have expenditure of the good since their consumption is zero;
>
> **iii) Binary/Dummy/Dichotomous Dependent Variable.** When Yis Y=0 or Y=1. For example, a consumer may or may no have a car/cable/computer/broadband service/ a house;
>
> **iv) Grouped Dependent variable.** A dummy variable that exist for M different groups so that Yj=pj is the percentage of agents which individual dependent variable is Yij=1 and Yj=1-pj is the percentage of agents which individual variable is Yij=0; j=1,..M; i=1,..Nj;
>
> **v) Polychotomous Dependent variables.** When Y= 1,2, ,,,K different values. For example; type of occupation: laywer (1); economist (2)…and so on;

This section will concentrate in the last three kinds of dependent variables.

2.1 Linear/Logit / Probit (Normit)/ Dummy Dependent Variables.

2.1.1 Linear Model.

$Yi= Xi.\beta + \varepsilon i=1; \varepsilon i= 1-Xi.\beta$
$Yi = Xi.\beta + \varepsilon i=0; \varepsilon i= -Xi. \beta$

$E(Yi/Xi)= f(1-Xi.\beta)=(Xi. \beta)$

$Var(Yi/Xi)=V(\varepsilon i/Xi)= (1-Xi. .\beta)^2. (Xi. .\beta)+ (Xi. .\beta)^2. (1-Xi. .\beta)$

$= (1-Xi. .\beta). (Xi. \beta).[(1-Xi. \beta+Xi. \beta]=(1-Xi.\beta). (Xi. \beta)= E(Yi/Xi).[1-E(Yi/Xi)];$

Since the variance is not homocedastic then LS is not a efficient estimator (i.e., the variance will not be minimum). The efficient estimator will be the Generalized Least Square estimator[6]. One simple form to obtain this latter estimator is by transforming the data to:

$Yi/ [Yie.(1-Yie)]^{1/2} = Xi/ [Yie.(1-Yie)]^{1/2} + \varepsilon i/[Yie.(1-Yie)]^{1/2}$

Where Yie is the estimator of the regression $Yei= Xi.\beta e$

The problem is that [Yie.(1-Yie)] could be negative and Yie could be greater than one.

In general, transforming the data to another set that satisfy the assumptions of the classical model is the best way to deal with the estimation procedure whenever the original data set does not satisfy the assumption discussed in the classical model. On the other hand, the standard LS estimator of the new transformed data set can be estimated to make inferences on the original model and data set.

2.1.2. Probit Model.

The model is transformed in :

$Y^*= X.\beta + \varepsilon$; $Yi=1$ if $Yi^* >0$; $Yi=1$ if $Yi^* \leq 0$;

Then $Prob (Yi=1)= Prob(Yi^* >0)= Prob. (\varepsilon >-Xi.\beta)= 1- F(\varepsilon < -Xi.\beta)$;;
And $Prob (Yi=0)= Prob (\varepsilon \leq -Xi.\beta)= F(\varepsilon i < -Xi.\beta)$;

[6] See any of the econometrics textbooks references.

$E(Yi/Xi)= 1-F(-Xi.\beta)$
$V(Yi/Xi)= F(-Xi.\beta). (1-F(Xi.\beta))$

The likelihood function is:

$$L= \Pi\, P(Yi)= \Pi\, [Prob\,(Yi=1)].\, \Pi\, [Prob\,(Yi=0)]= \prod_{Yi=1} [1- F(-Xi.\beta)].\, \prod_{Yi=0} F(-Xi.\,\beta).$$

To estimate β will depend upon the assumptions that is made about the cumulative distribution F(.).

Probit or Normit model assume that F(.) comes from a cumulative normal standard distribution. That is:

$$F(-Xi.\,\beta)= \int_{-\infty}^{-Xi.\beta} [2.\pi.]^{-1/2}.\, Exp\{(1/2)z^2$$

Note that $\sigma i^2 = V(Yi/Xi)= F(-Xi.\beta). (1-F(-Xi.\beta))$

Do knowing the parameters β we know also the variance

The Probit estimation is the MLE of the likelihood function L.

2.1.3 The Logit Model.

It is when $F(-Xi.\beta)= [\exp(-Xi.\beta)]. [1+ \exp(-Xi.\beta)]^{-1} =[1+ \exp(Xi.\beta)]^{-1}$

This distribitution is called the logistic distribution where $f(x)= \exp(x)/(1+\exp(x))^2$
$1- F(-Xi.\beta)= [\exp(Xi.\beta)]. [1+ \exp(Xi.\beta)]^{-1}$

The logit model also use the MLE of the parameters.

Note that the linear model assume that $F(-Xi\beta)=-Xi.\beta$

Let βl; βp; βli be the estimator of the logit, probit and linear model, then according to Amemiya (1981) to compare the three estimate we need to transform the estimates using the following relationships

$\beta p= (1.6)^{-1}. \beta l =(0.625)\, \beta l$;
$\beta li= (0.4). \beta p$ no constant; $\beta li= (0.4). \beta p + 0.5$ for the constant term;
$\beta li=(0.25)\, \beta l$ no constant; $\beta li= (0.25)\, \beta l + 0.5$ for the constant term

Example:

Linear Model

```
============================================================
              Dependent Variable: Y
              Method: Least Squares
          Date: 10/14/00   Time: 09:13
                Sample: 1950 2010
            Included observations: 61
============================================================
    Variable      CoefficientStd. Errort-Statistic  Prob.
============================================================
     C          0.172661   0.144148   1.197802   0.2358
     X          0.501948   0.227807   2.203386   0.0315
============================================================
R-squared            0.076030    Mean dependent var 0.459016
Adjusted R-squared   0.060370    S.D. dependent var 0.502453
S.E. of regression   0.487050    Akaike info criteri1.431339
Sum squared resid    13.99587    Schwarz criterion  1.500548
Log likelihood      -41.65584    F-statistic        4.854908
Durbin-Watson stat   0.774098    Prob(F-statistic)  0.031484
============================================================
```

Fuente: Eviews[7]

Probit Model

```
============================================================
              Dependent Variable: Y
           Method: ML - Binary Probit
          Date: 10/14/00   Time: 09:18
                Sample: 1950 2010
            Included observations: 61
       Convergence achieved after 3 iterations
    Covariance matrix computed using second derivatives
============================================================
    Variable      CoefficientStd. Errorz-Statistic  Prob.
============================================================
     C         -0.868882   0.395948  -2.194437   0.0282
     X          1.324741   0.616894   2.147437   0.0318
============================================================
Mean dependent var   0.459016    S.D. dependent var 0.502453
S.E. of regression   0.487082    Akaike info criteri1.366752
Sum squared resid    13.99770    Schwarz criterion  1.435961
Log likelihood      -39.68594    Hannan-Quinn criter1.393876
Restr. log likelihoo-42.07683    Avg. log likelihoo-0.650589
LR statistic (1 df)  4.781778    McFadden R-squared 0.056822
```

[7] The Eviews platform was extracted to perform the respective regressions.

Probit Model

```
==============================================================
                  Dependent Variable: Y
               Method: ML - Binary Probit
              Date: 10/14/00   Time: 09:18
                   Sample: 1950 2010
               Included observations: 61
         Convergence achieved after 3 iterations
    Covariance matrix computed using second derivatives
==============================================================
   Probability(LR stat) 0.028762
==============================================================
Obs with Dep=0            33     Total obs              61
Obs with Dep=1            28
==============================================================
```

Fuente: Eviews

Logit Model

```
==============================================================
                  Dependent Variable: Y
               Method: ML - Binary Logit
              Date: 10/14/00   Time: 10:08
                   Sample: 1950 2010
               Included observations: 61
         Convergence achieved after 3 iterations
    Covariance matrix computed using second derivatives
==============================================================
    Variable      CoefficientStd. Errorz-Statistic  Prob.
==============================================================
     C          -1.404377   0.657424  -2.136182   0.0327
     X           2.143307   1.021890   2.097394   0.0360
==============================================================
Mean dependent var   0.459016    S.D. dependent var 0.502453
S.E. of regression   0.487103    Akaike info criteri1.366810
Sum squared resid    13.99889    Schwarz criterion  1.436019
Log likelihood      -39.68770    Hannan-Quinn criter1.393934
Restr. log likelihoo-42.07683    Avg. log likelihoo-0.650618
LR statistic (1 df)  4.778254    McFadden R-squared 0.056780
Probability(LR stat) 0.028821
==============================================================
Obs with Dep=0            33     Total obs              61
Obs with Dep=1            28
==============================================================
```

Fuente: Eviews

βli =0.5919; βp =1.3247; βl= 2.1433; αli =0.1727; αp =-0.8699; αl= -1.4044;

βp= $(1.6)^{-1}$. βl =(0.625) βl; βli= (0.4). βp no constant; βli= (0.4). βp + 0.5 for the constant term; βli=(0.25) βl no constant; βli= (0.25) βl + 0.5 for the constant term

$\beta p'$= (0.625) βl= 1.3396; $\beta p''$=1.4797 so more or less they are the same number;
$\alpha p'$= -0.8778; $\alpha p''$= -0.8183; more or less the same number;

2.1.4 Hypotheses Testing.

Note that the logit/probit model the estimation method is MLE with all the asymptotic properties discussed in the former section. So the tests are similar to the ones employed either in the LS estimation or the test that comes from the Likelihood function. Eviews present most of the statistics of the LS estimation output (see above) and the likelihood ratio. As measure of fit, the Re^2 equivalent is defined by McFadden[8] (1974)

$$Rmfd^2 = 1- [\ln L(\beta_{MLEu})/ \ln L(\beta_{MLER})]$$

But $LR= 2. [\ln L(\beta_{MLEu})- \ln L(\beta_{MLER})]= 2. \ln L(\beta_{MLER}).(-1). Rmfd^2$

$$Rmfd^2= [-LR/2]. [\ln L(\beta_{MLER})]^{-1};$$

For the Probit example E-views yields the following results:

LR= 4.781778 ; $[\ln L(\beta_{MLER})]^{-1}$ =1/(-42.07683)= - 0.023766; $Rmfd^2$= 0.0568222.

2.2 Grouped Data.

This is the case where exist m groups of data and pj = nj/Nj is the probability that the event Yij=1 for i=1, ..nj has occurred wherein nj is the number of times that the event occurred in groupj and Nj is the number of individuals/units in the group j (and nj<Nj); and 1-pj is the probability that the event Yij=0 has occurred. If we have the total set of data it is possible to apply the former methods or the grouped data set discussed here. The specification models found in the literature of grouped data are:

> $pj= Xj.\beta + \varepsilon j$; the linear model [note that Xj is also grouped data]; j=1, m (number of groups)
> $E(\varepsilon j/Xj)=0$ or $E(pj)=\rho j$ [the true probability]; $V(\varepsilon j/Xj)= V(pj/Xj)= \rho j.(1-\rho j)/Nj$

[8] On Friday October 13, 2000; Daniel McFadden (from UC/Berkeley, 63 years old) and James Heckman (U.Chicago, 56 years old) were designated as the Nobel Price winners in economics because of their contribution of the models discussed here.

$lnpj = Xj.\beta + \varepsilon j$; is the log linear model;
$E(\varepsilon j/Xj)=0$; $V(\varepsilon j/Xj)= (1-\rho j)/(Nj.\ \rho j.)$;

$\ln[pj.(1-pj)^{-1}] = = Xj.\beta + \varepsilon j$; is the logit model
$E(\varepsilon j/Xj)=0$; $V(\varepsilon j/Xj)= [Nj.\ \rho j.(1-\rho j)]^{-1}$;

$G(pj)= Xj.\ \beta + \varepsilon j$; $G(pj)= F^{-1}(pj)$[9] and $pj= F(Xj.\beta)$; F is the cumulative Standard Normal Distribution; this is the normit/probit model
$E(\varepsilon j/Xj)=0$; $V(\varepsilon j/Xj)= [\rho j.(1-\rho j)]/[Nj.\ f([\rho j)^2]^{-1}$; f is the density of the standard normal distribution;

Note that $d\,F^{-1}(pj)/dpj= 1/f(pj)=dG/dpj$ then $dG= dpj/ f(pj)$

Transforming the data for each case using the square root of its respective variance and applying the LS estimator to this transformed data, we will have: i) the Minimum Chi square method for both the linear and log linear model; ii) and the minimum logit chi square method for the logit model. Note that in these models $pj \neq 0$ or 1. For cases where pj=0 or 1, Cox (1970) transformation is:

$\ln [pj+ (2.Nj)^{-1}]/[1-pj + (2.Nj)^{-1}]^{-1}$

The minimum normit chi square method is the LS of the transformed data using the square root of the variance of the probit model. (Maddala, 1983, section 2.8)

2.3 Polychotomous Unordered Dependent Variables.

Suppose now that Yj= 1, 2, ..g categories; let pij= the probability that the individual i has selected the j category, then the Likelihood function will be:

$$L= \prod_{i=1}^{N} f(Yi/Xi)= \prod_{i=1}^{N1} f(Yi1=1) \prod_{i=1}^{N2} f(Yi2=2)\ldots\ldots \prod_{i=1}^{Ng} f(Yig=G);\ \sum_{j=1}^{g} Nj=N$$

$$L = \prod_{i=1}^{N} [pi1]^{Yi1}.\ [pi2]^{Yi2} \ldots\ldots[pig]^{Yig} =\prod_{i=1}^{N} \prod_{j=1}^{G} [pij]^{Yij}$$

Where Yij=1 if the individual i fall in the category j and Yij=0 if it does not;

[9] En Eviews the command that yields this result is $F^{-1}(pj)=$ *@qnorm(pj).*

In order to get the estimation we need to assume some known distribution to estimate pij
Taking the logistic function then:

$$p_{ij} = [\exp(X_i.\beta_j)].\,[\,1 + \sum_{j=1}^{g-1} \exp(X_i..\beta_j)]^{-1};\quad p_{ig} = [\,1 + \sum_{j=1}^{g-1} \exp(X_i.\,\beta_j)]^{-1}$$

i=1, N; j=1, …G

The MLE of the βj is called the multinomial logit model. This Model is called unordered since the probabilities and responses are independents. The case of ordered (not so often used in economics practice) occurs when this probabilities and/or responses are dependent each other. Example of the former is the transportation vehicle: car, bus, train, airplane, etc. The fact an individual use a car does not depend upon that another individual has used car or any other vehicle. Further, an individual may use different kinds of transportation. An example of the latter is the effect of an insecticide on an insect. There are only three possibilities of the insect (after the effect of the insecticide): alive, moribund or death. Thus, the tree responses are related for the same insect.

2.4 Final Remarks.

The same interpretations of the LS model is apply also for limited dependent variables. Some differences do exist in the interpretation of the parameter β according to the model that it is used. For example, the linear model the parameter β measure the slope of each of the independent variables relative to the probability dependent variable. In the other models, you need to determine the slope or the elasticity case by case (see Judge and associates, 1988)

Hypotheses testing for the discrete dependent variables also follows the same techniques developed for the MLE and applying to the Logit, Probit, Chi Square (GLS = Generalized Least Squares) and Multinomial models (see Amemiya, 1981).

REFERENCES

Amemiya, T., 1981, "Qualitative Response Models: A Survey". Journal of Economic Literature, Vol. XIX, December, pp. 1483-1536.

Cox, D., 1970, Analysis of Binary Data. London, Methuen.

Dhrymes, P., 1994, "Limited Dependent Variables". ". In Handbook of Econometrics, vol. III, chapter 27.

Judge, G., R.Hill, E. Griffiths, H. Lupkepohl, T. Lee, 1985, The Theory and Practice of

Econometrics. Jonh Wiley and Sons.

McFadden, D., 1994, "Econometric Analyisis of Qualitative Response Models". In Handbook of Econometrics, vol. II, chapter 24.

McFadden, D., 1974, "Conditional Logit Analysis of Qualitative Choice Behavior". Frontiers in Econometrics. Eds. P. Zarembka, N.Y, Academic Press, pp. 105-142.

Maddala, G., 1983, Limited-Dependent and Qualitative Variables in Econometrics, Camdridge University Press.

3.TIME SERIES ANALYSIS

As in former chapters, this note presents a very brief glance of the statistical theory of Time Series. The random element of the data continues to be the key of all the statistical analysis. This case is applied to variables that evolve through time t wherever is the time frequency of the period. The span period of the time series sample is determined to be from 1 to T.

3.1 Basic Concepts.

Let y= $\{y1, \ldots. y_T\}$ a sample or sequence of a random variable which has as density function ft(yt); then yt [for t=1,….T] is called a random or stochastic process through time;

The former ***stochastic process is said to be strictly stationary if***:

$L(y1, \ldots., y_T)= L(y_{(1+k)}, \ldots\ldots y_{(T+k)})= L(y_{(1-k)}, \ldots\ldots y_{(T-k)})$

That is the joint distribution of the sample does not change regardless the span period +/- k. This implies that:

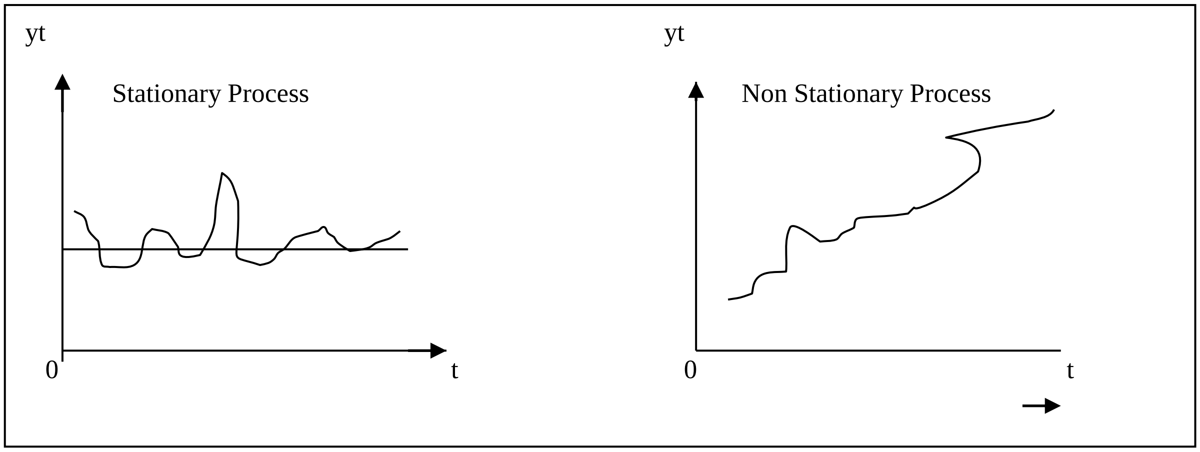
$E(yt)=\mu y$; $Var(yt)= \sigma y^2 < \infty$; $Cov(yt; yt+k)=Cov(yt-k; yt)= \gamma k$; t=1,…T; and $k\geq 0$

The expected value (or population mean); the variance and the covariance of any pair of the values of yt with a lag or forward time distance of k do not depend upon on time. The Covariance (Cov) in time series is called the auto-covariance.

A process is said to be weakly or covariance stationary if the process is strictly stationary and $\gamma k\neq 0$, i.e, the auto-covariance exist and is different from zero. A process is not stationary when the former conditions does not hold. Graphically speaking these processes look like the following figures:

Figure N° 4:

Stochastic process



A White Noise process is a strictly stationary process that satisfy the following conditions:

E(yt)= E(yt)=μy; Var(yt)= $\sigma y^2 < \infty$; Cov(yt; yt+k)=Cov(yt-k; yt)= γk=0. That is for any time series the yt , yt+k or yt-k are independent random variables and the joint density is:

$$L(y1,\ldots.y_T)= \prod_{t=1}^{T} f(yt)$$

A Simple Random Walk process is a non-stationary process such that applying the difference and one lag operator the process will change a white noise process. That is:

yt= yt-1 + εt; where εt is white noise stationary process

yt-yt-1= εt; let B called the lag operator such that B^kyt= yt-k and the difference operator Δ such that Δyt= yt –yt-1= yt-Byt= (1-B)yt; so a simple random walk is:

Δyt= (1-B)yt= εt;

Note (1-B)= Φ(B) is called the characteristic equation. The solutions that solve this polynomial equation are called roots. The real roots could be unit values; greater than unit values or lower than the unit values. The first case are called unit roots, when that happens the process is a random walk process or a unit root process. If the roots are higher than one (in absolute terms) the characteristic equation can not be inverted and the process will continue to be non-stationary. If the roots are lower than one (in absolute terms) then the characteristic equation could be inverted in $\Phi^{-1}(B)=\Theta(B)$ such that:

$$yt= \Theta(B)\varepsilon t;$$ (i.e., a Moving Average covariance Stationary process);

An non-stationary stochastic process homogeneous of degree h is a process where:

$\Delta^h yt = \varepsilon t$; and εt is a white noise stationary process;

3.2 ARIMA Processes.

Box and Jenkins (1976) developed a series of techniques in order to estimate standard non stationary process such as:

Moving average covariance stationary stochastic process of order q, MA(q): {yt} t=1, …T is one where:

$$yt= \alpha + \varepsilon t - \sum_{j=1}^{q} \theta j.\ \varepsilon t\text{-}j$$; εt a white noise process; $E(\varepsilon t)=0$ and finite variance σ^2

$E(yt)= \alpha$; $V(yt)= (1 + \sum_{j=1}^{q} \theta j^2).\ \sigma^2 < \infty$

$$yt=u + \varepsilon t -\sum_{j=1}^{q} \theta jB^j\varepsilon t= u + (1-\sum_{j=1}^{q}\theta jB^j)\varepsilon t =u +\Theta(B)\varepsilon t;\quad \Theta(B)= (1-\sum_{j=1}^{q}\theta jB^j)$$

MA(q) is an invertible process (and can be converted to an stationary process) if the roots of the equation $\Theta(B)$ are in absolute value greater than one or that the roots be outside of the circle of radio equal to one. This is called the unit circle. Then:

$\Phi(B).(yt-\alpha)= \varepsilon t$ is a white noise process and $\Theta^{-1}(B)= \Phi(B)$;

The auto-correlation function is the function that generates the auto-correlation coefficients of order zero. That is:

$\rho k= cov(yt, yt+k)/[V(yt).V(yt+k)]^{1/2}$; for k= -T…-1,.0,1, … T

Note that for a white noise process $\rho o=1$; $\rho k=0$

For a MA(q), it can be proved that $\rho k \neq 0$ for k= -q, …1,.0, 1, q; and $\rho k=0$ for k >/q/

The autoregresive non-stationary process yt of order p AR(p) is a process defined by:

$$yt= \delta + \sum_{i=1}^{p} \phi i.y_{t-i} + \varepsilon t,$$ where εt is a white noise process.

$$E(yt)= \delta/[1-\sum_{i=1}^{p}\phi i]; \quad \text{and} \quad / \sum_{i=1}^{p}\phi i/<1; \quad V(yt)= \sigma^2. [1- \sum_{i=1}^{p}\phi i^2]^{-1}; \sum_{i=1}^{p}\phi i^2 <1$$

$$(1- \sum_{i=1}^{p}\phi i.B^i)yt = \Phi(B).yt= \delta + \varepsilon t,$$

yt is covariance stationary if and only if there exist $\Phi(B)^{-1}$. That is the unit roots of the equation $\Phi(B)=0$ be outside of the unit circle. It can be proved that:

$\rho k \neq 0$ for k= -p, …1,.0, 1, p; and $/\rho k/ \rightarrow 0$ for k >/p/

The ARMA (p.q) non-stationary process of an Autoregressive process of order p and an Moving Average of order q is defined by:

$$yt=u + \varepsilon t -\sum_{j=1}^{q} \theta jB^j\varepsilon t + \sum_{j=1}^{p} \phi i.y_{t-i}$$; εt a white noise process with mean zero and finite variance

$$\Phi(B).yt = u + \Theta(B)\varepsilon t;$$

Note that the conditions for yt to be covariance stationary process are the same as the AR(p).
Finally the ARIMA (p,d,q) non stationary process is an ARMA(p,q) process for the variable Δ^d yt. That is:

$$\Delta^d yt=u + \varepsilon t -\sum_{j=1}^{q} \theta jB^j\varepsilon t + \sum_{j=1}^{p} \phi i.y_{t-i}$$; εt a white noise process with mean zero and finite variance

$$\Phi(B). \Delta^d yt = u + \Theta(B)\varepsilon t;$$

Once again, the conditions for yt to be covariance stationary process are the same as the AR(p).

Estimation of the ARIMA model previous identification (see below) can be done using, LS, GLS, MLE and non linear LS (detail in Box-Jenkins, 1976)

3.3 Spurious Regression and Stationary Process Tests

¿Why a stationary process is important?. Suppose we have the as a data the Annual Consumer Price Index that measure the level of price of an economy. In a period span of 50 years, it is found usually that this index goes up or have an increasing trend. The same we can say for the population of any medium or small country. So in general we could expect a high correlation between this two variables but without any meaningful explanation of that correlation. In this case, we say that correlation is spurious. If we have more than two variables of the same type we will have a spurious regression. From, large samples or asymptotically what we need is that all variables in the model come from stationary process so that we will not have spurious correlation. The former section provide some tools of how we can transform the data (but always taking account the theory behind the relationships) so that non-stationary be transformed stationary stochastic processes. Of course there are many procedures (methods) to do that, here it is summarized the most known methods.

3.3.1 Unit Root Tests

Suppose that we have two alternative processes for a stochastic variable:

Ho: yt= αo + yt-1 + εt; where εt is a white noise stationary process with zero mean and finite variance

H1: yt= αo + β. t + εt; where εt is a white noise stationary process with zero mean and finite variance
t=1, 2,..T

If Ho is accepted then the transformation that yields a stationary process is Δyt. If H1 is accepted then the transformation needed is yt-yet; where ye= αoe + βe. t, i.e., the equation is estimated by LS. The test that discriminate between Ho and H1 is called the Unit Root test and is based upon the following regression:

$$\Delta yt = \gamma o + \gamma .t+ (\alpha -1).yt-1 + \varepsilon t; = \gamma o= \alpha o. (1-\alpha); \gamma = \beta .(1-\alpha)$$

So if Ho': α=1 then Ho is true; if H1': α <1 then H1 is true. This is so called the Dickey-Fuller test (1979 and Fuller, 1976) or DF test. The Augmented DF test or ADF test assumes the existence auto correlation in the error term. The Ho could have a constant and a trend too and there exist tables for these tests. Another possibilities are found in Perron and Phillips unit roots tests (1988).

3.3.2 ARIMA Identification Tests.

In this case, what is needed is to identify the ARIMA (p, d, p) process and the to transform the variable to reduce to the white noise process.

To identify d, you may use a unit root test. Alternatively you may apply the difference operator "d" times to the original data and estimate the auto-correlation function. Thus, in the d time if the $\rho e_1=1$ and $\rho e_{1k}=0$ for k>1 (ρe is the estimator of ρ) then the transformed series Δ^d yt is a white noise stationary process. To test the former, there are to ways to do it. One is using the prob-value (i.e, 10% or less) for the individual statistical significance of the estimator of each auto-correlation coefficient (which is a t-student statistic of T-k degrees of freedom) or to use the statistic:

$$Q= T.\sum_{j=1}^{k} (\rho ej)^2$$; If Q >Q* then Ho: $\rho ej =0$; j >k is rejected; (Q is distributed as a Chi-Square with k degree of freedom)

The Degree of the MA Process (q) is also identified with the two previous tests. In this case for $\rho ej \neq 0$ for j=1, q; and $\rho ej = 0$ for j>q;

The Degree of the AR Process (p) using the partial correlation coefficients that are the coefficients ϕj of the AR process assuming p=1, 2, ..k. The statistical significance of each coefficient will determine the degree of the AR process. To estimate the partial autocorrelation coefficient that are the estimates of ϕj[10] assuming different values of p. Beginning with p=1, the estimate of ϕe_1 and then successively.

3.3.3 Filters and the Kalman Filters.

Filters are used in time series analysis to transform the original data to a smoother behavior of the variables in the data set than in the original data set. Sometimes, it means to transform a non-stationary series to a stationary series.

The Kalman Filters is a way to estimate state space representation of dynamic systems and it is used to evaluate the likelihood function and to forecast and smooth the unobserved state variables (See Hamilton ,1994). Another interpretation is that the model estimate coefficients that are random variables.

[10] There exist functions that relates the ϕj coefficients and the ρj coefficients (see Pindyck-Rubinfeld, 1981)

3.4 Cointegration and Causality Tests.

Spurious regression usually comes from variables (dependent and independent) that are non-stationary. However, there exist an special case where a set of variables despite to be a random walk non stationary processes still the regression does not yield spurious results. When this occurs, the set of variables is said to be co-integrated variables. The stochastic co-integration concept has also been used in the literature to define statistically the equilibrium of a set of variables. For instance, if the quantity demanded and supplied of a good are in equilibrium then their difference must be a stationary process that converges to zero. Formally, co-integration is defined by:

$\nu t = Y\eta - X\beta$; if Y and X are integrated processes of order 1 i.e, I(1) (meaning a random walk process) and νt is a stationary process then the variables Y and X are said to be co-integrated. Note that co-integration does not provide information on causality. There are several methods to test co-integration. The most known are using the Unit root test for νt, Y and X or the Johansen (1988, 19991).

Statistical causality test through time also have been defined and provided a test by Granger (1969). The test is based upon a model where the past or lagged independent variables are statistically significant to explain the dependent variable. The Granger causality test is the respective F test for the lagged coefficients of the independent variables.

3.5 VARs and Error Correction Models.

Two alternatives econometric methodologies have been developed in the context of time series. One by Sims (1980) and the other by Hendry and associates (1983, 1990 and Mizon, 1984).

Sims's methodology is a direct critic of large size macro econometrics models where identification is assumed according with the taste of the research. He argue that these models can be replaced for small ***parsimonious*** model where all the variables in the system are interrelated without imposing identifying restrictions. His approach extent the approach of the ARIMA models to a system of more than one dependent variable. Thus his estimate the ***Vector Auto regression*** Model or VAR defined by:

$$Yt = \alpha + \sum_{j=1}^{p} Yt\text{-}j.\Phi j + \varepsilon t;$$ where Yt is a vector of g endogenous variables

According to the assumption of the error term the VAR model could be estimates using LS, MLE or other methods (e.g., Johansen, 1988, 1991).

The Hendry (and associates) methodology is a third econometric methodology with respect to the analysis of time series. The methodology considers that the variables come from the theory and the dynamics is determined by the data. The notion of equilibrium is directly incorporated in the final error correction model (ECM). The usual specification is:

$\Delta yt = \alpha + \gamma. (yt\text{-}1 - \lambda. xt\text{-}1) + \gamma 1. \Delta xt + \varepsilon t;$

where $\gamma. (yt\text{-}1 - \lambda. xt\text{-}1)$ is the error correction term where it is assumed that xt and yt are co-integrated variables. It is possible to estimate the equation, estimating first λ regressing yt on xt the coefficient of xt is a estimated of first λ. Then estimate the former equation to estimate the rest of parameter. Note that λ is a adjustment parameter.

REFERENCES

Box, G., G. Jenkins, 1976, Time Series Analysis, Forecasting and Control. Golden Day, San Francisco.

Dickey, D., W. Fuller, 1979, " Distributions of the Estimates for Autoregressive Time Series with Unit Roots". *Journal of the American Statistical Association*, 74, June, pp 427-431.

Engle, R., C. Granger, 1987, "Co-Integration and Error Correction: Representation, Estimation and Testing". *Econometrica*, 55, pp. 251-276.

Fuller, W., 1976, Introduction To Statistical Time Series. John Wiley & Sons, N.Y.

Granger, C., 1969, "Investigating Causal Relations by Econometrics Models and Cross Spectral Methods". *Econometrica*, 37, pp. 425-438.

Granger, C., P. Newbold, 1974, "Spurious Regression in Econometrics". Journal of Econometrics, 2, pp. 111-120.

Hamilton, J., 1994, Time Series Analysis. Princeton University Press, New Jersey.

Harvey, A., 1989, Time Series Models. Oxford University Press.

Harvey, A., 1994, Forecasting Structural Time Series Models and The Kalman Filter. Cambridge University Press.

Hendry, D., J.F. Richard, 1983, "On the Formulation of Empirical Models in Dynamic Econometrics". *Journal of Econometrics*, 20, pp. 3-33.

Hendry, D., N. Ericsson, 1990, " An Econometric Analysis of U.K. Money Demand in Monetary Trends in the United States and the United Kingdom by M. Friedman and A. Schwartz". ". *American Economic Review*, No 81, pp. 8-38.

Intrilligator M., Z. Griliches, 1983, Handbook of Econometrics, Vol II, Part 5, Time Series Topics.

Johansen, S., 1988, "Statistical Analysis of Cointegration Vectors". *Journal of Economic Dynamic and Control*, 12, pp. 231-254.

Johansen, S., 1991, "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models". *Econometrica*, 59, pp. 1551-1580.

MacKinnon, J., y R. Davidson, 1993, Estimation and Inference in Econometrics. Oxford University Press.

Mizon, G., 1984, "The Encompassing Approach in Econometrics". In D.Hendry, K. Wallis, Econometrics and Quantitative Economics, pp. 135-172.

Perron, P., P. Phillips, 1988, " Testing for a Unit Roots in Times Series Regression". *Biometrica*, 75, June, pp. 335-346.

Pindyck, R., y D. Rubinfeld, 1981, Econometric and Models and Economic Forescasts. MacGraw Hill.

Sims, C, 1980, "Macroeconomics and Reality". *Econometrica*, 48, pp. 1-47.

4. PANEL DATA REGRESSION MODELS

The last topic in this note is the model using both cross-section and time series data. This set of data is called Pool Data or Panel Data. The linear model to be summarized here is:

$$[1] \qquad y_{it} = \beta_{oit} + \sum_{j=1}^{K} \beta_{jit} . x_{jit} + \varepsilon_{it}; \; i=1,\ldots N; \; t=1, \ldots T$$

The estimation methods of the model will depend upon the assumption of the stochastic term εit and the set of parameters βjit; j=0, K. The most known models/methods are the following:

The LS or Common coefficients model/method when:

[1.A.1] $\beta jit= \beta j$, for all i, t and j=0, …K;

$E(\varepsilon it/Xit)=0$; $E(\varepsilon. \varepsilon')= \sigma^2 [I_N \otimes I_T]$, where $\otimes$ is the Kronecker product

$A \otimes B= [aij.B]$ is kronecker product of two matrices, if A is of order mxn and B of order n'xm' then the kronecker product square matrix is of order mn'x nm'.

[1.A.2] $Var(\varepsilon it)= \sigma^2$ for any i,j and $cov(\varepsilon it, \varepsilon i't')=0$ for $i\neq i'$ and $t \neq t'$

Under those conditions the general model is transformed to:

[1]' $Y= X\beta + \varepsilon$;
where $Y=[y_{11}…..y_{N1}; y_{12}..y_{N2}; …. y_{NT}]'$;$\varepsilon=[\varepsilon_{11}.\varepsilon_{N1}; \varepsilon_{12}.. \varepsilon_{N2}; …. \varepsilon_{NT}]'$
$X=[x_0it; ….x_Kit]$; $xjit= [xj_{11…}xj_{N1}; xj_{12}.. xj_{N2}; …. xj_{NT}]'$

Y is a vector column of NTx1; X is a matrix of order NTxK; and ε is a vector column of order NTx1.

The individual or the time series model when:

[2.A.1] $\beta jit= \beta ji$, for all t,i and j=0, …K; [individual time series model for each i] or
$\beta jit= \beta jt$, for all i, t and j=0, …K; [individual cross section model for each t] and

$E(\varepsilon it/Xit)=0$; $E(\varepsilon. \varepsilon')= \sigma i^2 [I_N \otimes I_T]$, or
$E(\varepsilon it/Xit)=0$; $E(\varepsilon. \varepsilon')= \sigma t^2 [I_T \otimes I_N]$,

In this case the estimation are either individual time series model or cross section. That is:

[2] $Yi = Xi\beta i + \varepsilon i$; $Yi= [yi1….yi_T]'$; $\varepsilon i= [\varepsilon i1….\varepsilon i_T]'$; $Xi= [xio;…xi_K]$
$xij=[xij_1 …. xij_T]'$

Here it is estimated N individual and unrelated regressions. Alternatively:

[2]' $Yt = Xt\beta t + \varepsilon t$; $Yt= [yt1….yt_N]'$; $\varepsilon t= [\varepsilon t1….\varepsilon t_N]'$; $Xt= [xto;…xt_K]$
$xjt=[xjt_1 …. xjt_N]'$

In this case are estimated T regression models.

The Seemingly Unrelated Regression (SUR) Model it is the model that allows for interactions among individuals. The conditions are:

[3.A.1] $\beta jit = \beta ji$, for all t,i and j=0, …K; [individual time series model for each i]

$E(\varepsilon it/Xit)=0$; $E(\varepsilon.\ \varepsilon') = [E(\varepsilon it.\ \varepsilon js')] = [\sigma ij\ I_T]$, for t=s and [0] for $t \neq s$
For i,j=1, ..N and s,t=1, …T

The model then could be written as N equations:

[3] $Yi = Xi.\ \beta i + \varepsilon i$; i=1, ..N and $Yi=[yi1 \ldots yi_T]'$; $\varepsilon i = [\varepsilon i1 \ldots \varepsilon i_T]'$; $Xi = [xio; \ldots xi_K]$

That can be estimated simultaneously using the GLS of the unified equation system:

[3]' $Y = X.\ \beta + \varepsilon$; $Y=[Y_1 \ldots Y_N]'$; $\varepsilon = [\varepsilon 1 \ldots \varepsilon_N]'$; $X = \begin{pmatrix} X_1 & 0 & & 0 \\ & & & \\ 0 & & Xi & \\ 0 & & & X_N \end{pmatrix}$
$\beta=[\beta 1 \ldots \beta_N]'$

The Matrix of variance and covariance is $E(\varepsilon.\varepsilon') = \Sigma \otimes I_N$ and $\Sigma=[\sigma ij]$ and square matrix of order n. Note that a estimator of $\sigma eij = \varepsilon ei'.\varepsilon ej/(T-K)$ or $(\varepsilon ei'.\varepsilon ej)/T$; where $.\varepsilon ej$ are the error estimator of the individual regressions.

The former three cases are special cases of the Panel data model in the sense that the available and standard techniques are applied to those models. The next three Panel data models are the one that usually has been dealt with in the literature.

4.1 The Covariance Model or the Fixed Effect and Fixed Coefficient Models

The model in this case is written as:

[4] $$yit = \beta oi + \sum_{j=1}^{K} \beta j.xjit + \varepsilon it;\ i=1,\ldots N;\ t=1, \ldots T$$

for the individual intercept fixed model or

$$y_{it} = \beta_{ot} + \sum_{j=1}^{K} \beta_j x_{jit} + \varepsilon_{it}; \; i=1,\ldots N; \; t=1, \ldots T$$

for the time intercept fixed model. In both cases the assumptions on the error stochastic term are the same as the LS method above.

The estimation procedure is simple:

First, either model is transformed using the individual or time means of the variables so that the data transformed in deviations will eliminate the intercept (i.e., the individual or time fixed effect);

Second, estimate the LS or GLS depending upon the assumptions on the error term and find the estimators for the fixed coefficients;

Third, the individual (or time) effect is computed as:

$$\beta_{eoi} = y_{mi} - \sum_{j=1}^{K} \beta_{ej} . x_{mji};$$ y_{mi} and x_{mji} are the sample mean (across time) of each variable or

$$\beta_{eot} = y_{mt} - \sum_{j=1}^{K} \beta_{ej} . x_{mji};$$ y_{mi} and x_{mji} are the sample mean (across individuals of each variable);

To test Ho: $\beta_{oi}=\beta_{os}$ for some particular pair of individual (or time periods)
Then you use a t of NT-(N+K) degree of freedom where

$$t = [\beta_{eoi} - \beta_{eos}]/[V([\beta_{eoi} - \beta_{eos})]^{1/2};$$

To test Ho: $\beta_{oi}=\beta_o$ for all i=1, N and F statistics is used with N-1 degrees in the numerator and NT-(N+K) degrees of freedom in the denominator and where:

$$F = \{[\varepsilon e_R' \, \varepsilon e_R - \varepsilon e_U' \, \varepsilon e_U]/(N-1)\} / [(\varepsilon e_U' \, \varepsilon e_U)/(NT-(N+K))$$

The restricted model is the model assuming Ho i.e., no individual (time) fixed effects and the unrestricted model is assuming the individual (time) fixed effects.

4.2 The Error Component or Random Error Models

The Panel data model is written as:

[5] $Y= X.\beta + \varepsilon$; $\varepsilon= \mu i \otimes l_T + \nu$; Y is a column vector of order NTx1; X is a matrix of order NTxK; μi is a column vector of order Nx1; l_T is the column vector of ones of order Tx1 and ν is column vector of order NTx1. There are two main differences between this model and former one. One is the fact that the individual (could also be the time) fixed effect no longer is fixed but rather is a random variable. The other is that the effect is incorporated as unknown effect or stochastic effect of factors although individuals that are omitted from the true specification and that in average their effects cancel out. As in all the cases, the estimation procedure of this model will depend upon the assumptions on the stochastic term. The usual set of assumptions found in the Panel data literature is:

[5.A.1] $E(\varepsilon/X)=0$ or $E(\mu i)=0$; $E(\nu)=0$;

$V(\varepsilon)=E(\varepsilon.\varepsilon')= \sigma_\mu^2\, I_N \otimes (l.\ l') + \sigma_\nu^2.\ I_N \otimes I_T$;

$E(\mu i.\ \mu i') = \sigma_\mu^2\, I_N$; $V(\nu it)= \sigma_\nu^2$ for any i,j and $cov(\nu it, \nu i't')=0$ for $i \neq i'$ and $t \neq t'$

Let's define:

[D1] (yit-ymi) = Wy the within group of the y variable (or the y variable in deviations from its mean of each group i)

(xjit-xmji)= Wxj the within group of xj variable;

(ymi-ym)= By the between group of the y variable;

(xmji- xmj)= Bxj the between group of the xj variable;

(yit-ym)= Ty the total group of the y variable;

(xjit-xmj)= Txj the total group of the xj variable; and

Ty= Wy+ By; Txj= Wxj + Bxj;

Let εedi= Ydi-Yedi be the column vector Tx1 of the error term from the I group regression where the data is terms of deviations, then:

$E(\varepsilon edi)= E(\varepsilon di)= E(\nu i) =0$ under the assumptions of the model then:

The LS of the variance $E(V(\nu i))= \sigma_\nu^2$ is

N

$$[5.R.1]\ \sigma e_{\nu}^{2} = \sum_{i=1} (\varepsilon edi'\varepsilon edi)/[NT-(N+K)]$$

The MLE of the same variance is:

$$[5.R.1]'\ \sigma e'_{\nu}{}^{2} = \sum_{i=1}^{N} (\varepsilon edi'\varepsilon edi)/(NT)$$

Note these error terms come from the within regression model: That is:

$Wy = Wx.\beta + W\varepsilon$

Let now εe the estimator of the error term of the original model (which is also equal to the estimator of the error of the total group set of variables). The expected value of the LS estimator of the variance of this error term is:

$E[(\varepsilon e' \varepsilon e)/(NT-(N+K))] = \sigma^2 = \sigma_{\nu}^2 + \sigma_{\mu}^2$

Then $\sigma e_{\mu}^2 = \sigma e^2 - \sigma e_{\nu}^2$; the MLE estimator of σ_{μ}^2 is

$[5.R.2]\ \sigma e_{\mu}^2 = \sigma e'^2 - \sigma e'_{\nu}{}^2$;

using the respective MLE. Once those two variances are known then apply the GLS to the original equation to re-estimate β.

4.3 The Random Coefficient Model.

This model takes the following form:

$$[6]\qquad y_{it} = \beta_{oi} + \sum_{j=1}^{K} \beta_j . x_{jit} + \nu_{it};\ i=1,\ldots N;\ t=1,\ \ldots T$$

[6.A.1] $\beta_{oi} = \beta_o + \mu_i$; where the intercept coefficient now is random due to the error term of the coefficient μ_i.

Under the same assumptions between both error terms as the previous model, the estimation procedure is as follows:

First, estimate either by LS or MLE the within group N equations. Using the error terms of these N equations estimate the variance of ν_{it}, σe_{ν}^2 (see equations [5.R.1] or [5.R.1]');

Second, use the whole system of equation group as one equation using the between groups transformation. That is:

$$[6.1] \quad ymi = \mu i + \sum_{j=1}^{K} \beta j.xjit + \nu im; \; i=1,\ldots N;$$

$$[6.2] \quad ym = \mu m + \sum_{j=1}^{K} \beta j.xmj + \beta o + \nu m;$$

$$[6]' \quad By = Bx.\beta + (\mu i - \mu m) + (\nu im - \nu m) = Bx.\beta + \varepsilon_B \; ;$$

$$[6]'' \quad \text{Where } V(\varepsilon_B i) = V(\mu i - \mu m) + V(\nu im - \nu m) = T.\sigma_\mu^2 + \sigma_\nu^2 = .\sigma_B^2$$

The estimation of the error term of the between group equation [6]' is used to estimate σe_B^2 and from [6]'' to estimate the variance σ_μ^2;

Third using these estimates of the variance we apply GLE to the original model in [6] and estimate the parameters. Note that equation [6.2] can be used to estimate βo.

In general the statistical and asymptotic properties of the estimator of the model considered here, will depend upon the size of N, T and K and the plausibility of the assumptions of the error terms.

4.4 Unit Roots, Co-Integration,VARs and Causality in Panel Data.

These issues are relative new using Panel data. The main contributors are: Levin and Lin (1992, 1993), Holtz-Eakin, Newey and Rosen (1988), and Pedroni (1995, 1997). In 1999 in the Oxford Bulletin of Economics and Statistics was published an special issue in these topics (See, Banerjee, 1999).

The Basic Model for the Unit Root test is:

$$[1] \quad \Delta yit = \alpha i + \delta i.\, Time + \theta t + \rho i.yit\text{-}1 + \varepsilon it; \; i=1,..N; \; t=1\ldots T$$

where Time is a variable taking values from 1 to T. This model allows for fixed effects across units and time plus the trend variable. The Null Hypothesis is

Ho: $\rho i = 0$; H1: $\rho i = \rho < 0$; $T.N^{1/2}\rho e \Rightarrow N(0,2)$ when $T, N \to \infty$ and $t_\rho \Rightarrow N(0,1)$;

If $\theta t=\alpha i=\delta i=0$; then Ho: $\rho =0$; and $\theta t=\alpha i=\delta i=0$; H1: $\rho <0$; and coefficients unrestricted, the statistic is:

$T.N^{1/2}\rho e + 3.N^{1/2} \Rightarrow N(0,10.2)$ and $(1.25)^{1/2} t_\rho + (1.875.N)^{1/2} \Rightarrow N(0,645/112)$

When $N^{1/2} /T \rightarrow 0$ and $T, N \rightarrow \infty$

For each change of the assumptions the test is different. See more models test in Banerjee (1999).

Pedroni (1999) summarize the main Panel Data ***co-integrated models*** discussed in the literature and present also the respective tables for the statistics. The basic model is:

$$[2] \qquad yit= \alpha i + \delta i. Time + \sum_{j=1}^{M}\beta j.xjit + \varepsilon it; \quad i=1,..N; t=1\ldots T$$

where M is the number of exogenous variables. The test is to identify if there exist co-integration between yit and xjit. Most of the parametric statistics[11] are based on the estimation of the errors of equation [2]. That is:

$$[3] \qquad \varepsilon eit= \gamma i\, \varepsilon eit\text{-}1 + \sum_{k=1}^{K} \gamma ik.\Delta\, \varepsilon eit\text{-}k + \mu it;$$

The Ho= $\gamma i= 1$; H1: $\gamma i= \gamma <1$; (for the within dimension statistics) and H1: $\gamma i < 1$ for the between dimension statistics. If Ho is accepted then there is no co-integration and if Ho is rejected then there will be co-integration. Note under Ho, the estimated error term is a random walk process, which it should not be if variable y and x are co-integrated, under co-integration the error term should be a stationary process. The formulas for the different statistics (seven in total) are found in Pedroni (1999, Table No 1, pp 660-661).

The VAR and causality tests with Panel Data were originated in the work of Holtz-Newey-Rosen (1988). The original model that these author deal with is:

$$[4] \qquad yit= \alpha t + \sum_{j=1}^{M}\delta jt.xit\text{-}j + \sum_{j=1}^{M}\alpha jt.yit\text{-}j + \psi t.fi + \varepsilon it; \quad i=1,..N; t=1\ldots T$$

[11] Nonparametric techniques have not been described in these notes. In these techniques no assumptions are made on the existence of parameters of the underlying distributions.

This is a random effect models where $\psi t.fi$ is not observable and fixed time effect αt. The authors transform this model to:

$$[5] \qquad yit = at + \sum_{j=1}^{M+1} djt.xit\text{-}j + \sum_{j=1}^{M+1} cjt.yit\text{-}j + vit; \quad i=1,..N;\ t=1\ldots T$$

Where [5] is obtained differencing equation [4]; to estimate [5] it is required that $T \geq m+3$. Under standard assumptions on the error term these authors propose a GLS estimator. Causality test is found, making hypothesis testing on the set of coefficient djt.

Since the work of Romer (1986), Barro (1991) and Temple (1999) among others have applied panel data extensively on growth equation models. Criticism to this line of research are found in Srinivasan (1993); Levin and associates (1991, 1992, 1993) and Maddala (1999).

References

Banerjee, A., 1999, "Panel Data Unit Roots and Cointegration: An Overview". *Oxford Bulletin of Economics and Statistics*; 61, Special Issue Nov. , pp. 607-29.

Barro, R., 1991, "Economic Growth in a Cross Section of Countries". *Quarterly Journal of Economics*, vol 106, No 2, May, pp. 407-443.

Chamberlain, G., 1994, "Panel Data". In Handbook of Econometrics, vol. II, chapter 22.

Holtz-Eakin, D., W. Newey, H. Rosen, 1988, "Estimating Vector Autoregressions with Panel Data". *Econometrica*, 56, pp. 1371-1395.

Hsiao, C., 1986, The Analysis of Panel Data, Cambridge University Press.

Levine R., D. Renelt, 1991, "Cross Country Studies of Growth and Policy: Methodological, Conceptual and Statistical Problems", WPS 608, World Bank.

Levin, A., Lin, Chien Fu, 1993, "Unit Root Tests in Panel Data: New Results". University of California, San Diego Department of Economics Working Paper: 93-56, December.

Levin, A., Lin, Chien Fu, 1992, "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties". University of California, San Diego Department of Economics Working Paper: 92-93.

Levine, R., D. Renelt, 1992, "A Sensitivity Analysis of Cross-Country Growth Regression", *American Economic Review*, 84, pp. 426-430.

Levine, R., S. Zeros, 1993, "What we have Learned about Policy and Growth from Cross-Country Regressions", *American Economic Review*, 84, pp. 426-430, May.

Maddala, G., 1993, The Econometrics of Panel Data , Elgar Reference Collection Series, vol. No 1 and 2.

Maddala, G., 1999, " On the Use of Panel Data Methods with Cross Country Data". *Annales d'Economie et de Statistique*, Sept-Dec.

Matyas, L., P. Sevestre, 1992, The Econometrics of Panel Data: Handbook of Theory and Applications, Kluwer Academic Publishers.

Pedroni, P., 1999, "Critical Values for Co-Integration Tests in Heterogeneous Panels with Multiple Regressors". *Oxford Bulletin of Economics and Statistics*; 61, Special Issue Nov. , pp. 653-670.

Pedroni, P., 1997, "Panel Co-integration: Asymptotic and Finite Sample Properties of Pooled Time Series with an Application to the PPP Hypothesis, New Results". Indiana University Working Paper.

Pedroni, P., 1995, "Panel Co-integration: Asymptotic and Finite Sample Properties of Pooled Time Series with an Application to the PPP Hypothesis". Indiana University Working Paper, No 95-013.

Romer, P., 1986, "Increasing Returns and Long-Run Economic Growth". *Journal of Political Economy*.

Srinivasan, T.N., 1993, "Long Run Growth Theories and Empiric: Anything New". NBER, 4th annual East Asian Seminar in Economics, June 17-19.

Temple., J., 1999, "The New Growth Evidence". *Journal of Economic Literature*, vol. XXXVII, No 1, March, pp.112- 156.

REFERENCES

1. Books

Amemiya, T., 1985, Advanced Econometrics. Harvard University Press.

Bard, Y., 1974, Nonlinear Parameter Estimation. Academic Press.

Bates, C., D. Watts, 1988, Nonlinear Regression Analysis and Its Applications. John Wiley and Sons.

Box, G., G. Jenkins, 1976, Times Series Analysis, Forecasting and Control. Holden Day.

Charemza, W., y D. Deadman, 1992, New Directions in Econometric Practice: General To Specific Modelling, Cointegration and Vector Regression. Edward Elgar Publishing Limited.

Cook, Thomas; D. Campbell, 1979, Quasi-Experimentation: Design & Analysis Issues for Fields Settings, Houghton Mifflin Co.

Dhrymes, P. 1989, Econometric Theory. Jonh Wiley and Sons.

Encaoua, David, 1999, The economics and econometrics of innovation , Boston : Kluwer Academic Publishers.

Engel, R., B. Yoo, 1991, "Cointegrated Economic Time Series: A Survey with New Results". En Engel-Granger eds, Long Run Economic Relationships: Readings in Cointegration.. Oxford University Press.

Engle, R., C. Granger, 1991, Long Run Economic Relationship: Readings in Cointegration. Oxford University Press.

Fomby, Thomas B. ; Hill, R. Carter, 2000, Applying Kernel and Nonparametric Estimation to economic topics.
Goldberger, A, 1970, Econometric Theory.

Granger, C., P. Newbold, 1986, Forecasting Time Series. Academic Press.

Harvey, A., 1990, Time Series Models. Oxford University Press.

Harvey, A., 1994, Forecasting Structural Time Series Models and The Kalman Filter. Cambridge University Press.

Intriligator, M. , 1978, Econometric Model Techniques and Applications. Prentice Hall.

Intriligator, M., Z. Griliches, 1984, Handbook of Econometrics, Vol I, II, III. North Holland, Amsterdam.

James, Lawrence, S. Mulaik, J. Brett, 1982, Causal Analysis: Assumptions, Models and Data, Sage Publications.

Johnston, J., 1980, Econometric Methods.

Jorgenson, Dale, 2000, Econometric modeling of producer behavior , MIT Press.

Baltagi, Badi H., 2000, A companion to theoretical econometrics. Malden : Blackwell Publishers,

Judge, G., R.Hill, E. Griffiths, H. Lupkepohl, T. Lee, 1988, The Theory and Practice of

Kendall M., A. Stuart, 1961, The Advanced Theory of Statistics, Vol I, II C. Griffen and Co. Ltd.

Kmenta, 1980, Econometrics.

MacKinnon, J., y R. Davidson, 1993, Estimation and Inference in Econometrics. Oxford University Press.

Maddala, 1981,Econometrics.

Malinvaud, E., 1970, Statistical Methods of Econometrics. North Holland.

Pindyck, R., y D. Rubinfeld, 1981, Econometric and Models and Economic Forescasts. MacGraw Hill.

Rao, C., 1973, Linear Statistical Inference and Its Applications. John Wiley and Sons.

Rassel, Gary; E. O'Sullivan, 1995, Research Methods for Public Administrators, 2nd Edition, Longman Publishers.

Rigby, Paul, 1965, Conceptual Foundations for Business Research, John Wiley and Sons, Inc.

Sheth, Jagdish, 1977, Multivariate Methods for Market and Survey Research, American Marketing Association.

Theil, H., 1980, Principles of Econometrics. John Wiley and Sons.

Wentz, Walter, 1972, Marketing Research: Management Methods, Harper and Row Publishers.

Winkelmann, Rainer, 2000, .Econometric analysis of count data , New York : Springer

Wright, Sonia, 1979, Quantitative Methods and Statistics: A Guide to Social Research, Sage Publications.

Zellner, A., 1971, An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons.

2. Journals and Working Papers

Campbell, J. P. Perron, 1991, "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots". NBER Macroeconomics Conference, Marzo.

Dickey, D., W. Bell, R. Miller, 1986, "Unit Roots in Time Series Models: Tests and Implications". *The American Statistician*, 40, pp. 12-26.

Engel R., C. Granger, 1987, "Cointegration and Error Correction: Representation, Estimation and Testing". *Econometrica*, 55, pp. 391-407.

Engel R., M. Rostchild, 1992, Special Issue on ARCH models in Finance. *Journal of Econometrics*, 52, No 1/2.

Engel, R., 1982, "Autorregresive Conditional Heterokesdaticity with Estimates of the Variance of UK Inflation" (ARCH model). *Econometrica*, 50, pp 987-1007.

Engel, R., D. Hendry, J. Richard, 1983, "Exogeneity". *Econometrica*, 51, pp. 277-304.

Granger C., 1981, "Some Properties of Time Series data and their use in Econometric Model and Specification". *Journal of Econometrics*, 16, pp. 121-30.

Granger, C., 1969, "Investigating Causal Regressions by Econometrics Models and Cross-Spectral Methods". *Econometrica*, 37, pp. 424-38.

Granger, C., P. Newbold, 1974, "Spurious Regressions in Econometrics". *Journal of Econometrics*, 2, pp. 111-20.

Haussman, J., W. Taylor, "Identification in Linear Simultaneous Equations Models with Covariances Restrictions: An Instrumental Variables Interpretation". *Econometrica*, 51, 1527-49.

Hendry, D., 1976, "The Structure of Simultaneous Equations Estimators". *Journal of Econometrics*, 4, pp. 51-88

Hendry, D., 1986, " Econometric Modelling with Cointegrated Variables: An Overview". *Oxford Bulletin of Economics and Statisctics*, 48, pp. 201-12.

Hu, Luojia, 2000, "Estimating a censored dynamic panel data model with an application to earnings dynamics". Working paper, Industrial Relations Section, Princeton University, No.435.

Hylleberg., S., R. Engel; C. Granger; B. Yoo, 1990, "Seasonal Integration and Cointegration". *Journal of Econometrics*, 44, pp. 215-238.

Leamer, E., 1987, "Error in Variables in Linear Systems". *Econometrica*, 55, pp. 893-909.

Lee, L., 1978, " Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variable". *International Economic Review*, 19, pp. 415-33.

Lee, L., 1981, "Simultaneous Equations Models with Discrete and Censored Variables". En Manski-MacFadden, eds, Structural Analysis of Discrete Data with Econometric Applications, MIT Press.

Manski, Charles F. ; Goldberger, Arthur Stanley, 2000, "Principles of econometrics". *Journal of Econometrics*, v. 95, no. 2.

Morales, Roberto Antonio, 2000, "Measuring the risk of investment in Latin America's emerging markets". Virginia Polytechnic Institute and State University,

Sims, C., 1972, "Money, Income and Causality". *American Economic Review*, 62, pp. 540-52.

Sims, C., 1980, "Macroeconmics and Reality" (VAR models). *Econometrica*, 48, pp. 1-48.

American Economic Review
Econometric
Economic Journal
Journal of Applied Econometrics
Journal of Economic History

Journal of Economic Literature
Journal of Economic Abstracts
Journal of Economic Perspectives
Journal of Industrial Economics
Journal of Money, Credit and Banking
Journal of Political Economy
Quarterly Journal of Economics
Review of Economic Studies
Review of Economics and Statistics
Econlit (on line Data base)